

UNIVERSIDADE FEDERAL DO PARANÁ

OLIVER KOLOSSOSKI

ALGORITMOS HÍBRIDOS PROXIMAIS EXTRAGRADIENTES PARA OS
PROBLEMAS DE PONTO DE SELA E EQUILÍBRIO DE NASH

CURITIBA

2016

OLIVER KOLOSSOSKI

ALGORITMOS HÍBRIDOS PROXIMAIS EXTRAGRADIENTES PARA OS
PROBLEMAS DE PONTO DE SELA E EQUILÍBRIO DE NASH

Projeto de Tese apresentado ao Programa de Pós-Graduação em Matemática, Área de Concentração em Otimização, do Departamento de Matemática, Setor de Ciências Exatas, como parte obrigatória das exigências para a obtenção do Título de Doutor.

Orientador: LUIZ CARLOS MATIOLI

Coorientador: RENATO MONTEIRO

CURITIBA

2016



MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
Setor CIÊNCIAS EXATAS
Programa de Pós Graduação em MATEMÁTICA
Código CAPES: 40001016041P1

ATA Nº 011

ATA DE SESSÃO PÚBLICA DE DEFESA DE TESE PARA A OBTENÇÃO DO GRAU DE DOUTOR EM MATEMÁTICA

No dia dois de Setembro de dois mil e dezesseis às 14:00 horas, na sala Salão Nobre, Centro Politécnico, do Setor de CIÊNCIAS EXATAS da Universidade Federal do Paraná, foram instalados os trabalhos de arguição do doutorando OLIVER KOLOSSOSKI para a Defesa Pública de sua Tese intitulada: "ALGORITMOS HÍBRIDOS PROXIMAIS EXTRAGRADIENTES PARA OS PROBLEMAS DE PONTO DE SELA E EQUILÍBRIO DE NASH". A Banca Examinadora, designada pelo Colegiado do Programa de Pós-Graduação em MATEMÁTICA da Universidade Federal do Paraná, foi constituída pelos seguintes Membros: LUIZ CARLOS MATIOLI (UFPR), BENAR FUX SVAITER (IMPA), LUCAS GARCIA PEDROSO (UFPR), LUCELINA BATISTA DOS SANTOS (UFPR), MA TO FU (USP/SC). Dando início à sessão, a presidência passou a palavra ao discente, para que o mesmo expusesse seu trabalho aos presentes. Em seguida, a presidência passou a palavra a cada um dos Examinadores, para suas respectivas arguições. O aluno respondeu a cada um dos arguidores. A presidência retomou a palavra para suas considerações finais e, depois, solicitou que os presentes e o doutorando deixassem a sala. A Banca Examinadora, então, reuniu-se sigilosamente e, após a discussão de suas avaliações, decidiu-se pela APROVAÇÃO do aluno. O doutorando foi convidado a ingressar novamente na sala, bem como os demais assistentes, após o que a presidência fez a leitura do Parecer da Banca Examinadora. Nada mais havendo a tratar a presidência deu por encerrada a sessão, da qual eu, LUIZ CARLOS MATIOLI, lavrei a presente ata, que vai assinada por mim e pelos membros da Comissão Examinadora.

Curitiba, 02 de Setembro de 2016.


LUIZ CARLOS MATIOLI

Presidente da Banca Examinadora (UFPR)


BENAR FUX SVAITER

Avaliador Externo (IMPA)


LUCAS GARCIA PEDROSO

Avaliador Interno (UFPR)


LUCELINA BATISTA DOS SANTOS

Avaliador Interno (UFPR)


MA TO FU

Avaliador Externo (USP/SC)



MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
Setor CIÊNCIAS EXATAS
Programa de Pós Graduação em MATEMÁTICA
Código CAPES: 40001016041P1

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em MATEMÁTICA da Universidade Federal do Paraná foram convocados para realizar a arguição da Tese de Doutorado de **OLIVER KOLOSSOSKI**, intitulada: "**ALGORITMOS HÍBRIDOS PROXIMAIS EXTRAGRADIENTES PARA OS PROBLEMAS DE PONTO DE SELA E EQUILÍBRIO DE NASH**", após terem inquirido o aluno e realizado a avaliação do trabalho, são de parecer pela sua APROVAÇÃO.

Curitiba, 02 de Setembro de 2016.

LUIZ CARLOS MATIOLI
Presidente da Banca Examinadora (UFPR)

BENAR FUX SVAITE
Avaliador Externo (IMPA)

LUCAS GARCIA PEDROSO
Avaliador Interno (UFPR)

LUCELINA BATISTA DOS SANTOS
Avaliador Interno (UFPR)

MA TO FU
Avaliador Externo (USP/SC)

*Este trabalho é dedicado a meus pais, Fábio Mandira, Bruno Lago, Arager,
Abarges, Acarlton, à Metalstorm.net e a todos que fazem um papel
importante na minha vida.*

AGRADECIMENTOS

Agradeço principalmente ao meu orientador, professor Doutor Luiz Carlos Matioli, pelo constante acompanhamento do começo ao fim deste trabalho. Também ao Professor Doutor Renato Monteiro, da Georgia Institute Of Technology pelo acompanhamento em 2014-2015.

À Universidade Federal do Paraná e ao Programa de Pós-Graduação em Matemática, pela oportunidade de cursar o Doutorado e ao grupo docente do PPGM, pelos valiosos ensinamentos transmitidos. E também à CAPES pelo apoio financeiro.

Agradeço a Deus, à minha família e amigos pelo constante suporte.

RESUMO

Neste trabalho são descritos métodos para determinar uma solução (aproximada) para os problemas de ponto-de-sela (PS) e equilíbrio de Nash. Os algoritmos são instâncias especiais do método híbrido extragradiente proximal introduzido por Svaiter e Solodov [Solodov; Svaiter, 2000] onde os sub-problemas de inclusão são resolvidos com o uso de um método de gradiente acelerado. Os métodos propostos generalizam o algoritmo acelerado de [He; Monteiro, 2014] das seguintes maneiras: a) em uma generalização os problemas considerados são problemas PS gerais ao invés de problemas PS com estrutura bilinear; b) em outra generalização o algoritmo é baseado em distâncias de Bregman ao invés da distância Euclidiana; c) em outra generalização o problema considerado é o de equilíbrio de Nash ao invés do problema de ponto-de-sela. Assim como no método de He e Monteiro, os métodos propostos têm a vantagem de que qualquer escolha de escalar para o tamanho do passo pode ser utilizada. Ainda, no contexto de problemas de ponto-de-sela, para certa escolha do tamanho do passo pode-se obter uma complexidade ótima para o método. Resultados computacionais ilustram a performance dos métodos em comparação com o método de suavização de Nesterov [Nesterov, 2005].

Palavras-chaves: programação convexa, complexidade, convergência ergódica, operador monótono maximal, método híbrido extragradiente proximal, método de gradiente acelerado, problema de ponto de sela, problema de equilíbrio de Nash, distância de Bregman.

ABSTRACT

In this work we describe methods to find an (approximate) solution for the saddle-point (SP) and Nash equilibrium problems. The algorithms are special instances of a hybrid extragradient proximal method introduced by Svaiter and Solodov [Solodov; Svaiter, 2000] where the inclusion sub-problems are solved using an accelerated gradient method. The proposed methods generalize the accelerated algorithm of [He; Monteiro, 2014] in the following ways: a) in a generalization, the considered problems are general SP problems instead of SP problems with a bilinear structure; b) in other generalization, the algorithm is based on Bregman distances rather than the Euclidian one; c) in other generalization, the considered problem is the Nash equilibrium problem instead of the saddle-point. As in He and Monteiro's method, the proposed methods have the advantage that any scalar choice for the stepsize can be used. Also, for the saddle-point problems, a certain choice for the stepsize can yield an optimal complexity for the method. Computational results show the performance of the methods in comparison with Nesterov's suavization scheme [Nesterov, 2005].

Key-words: convex programming, complexity, ergodic convergence, maximal monotone operator, hybrid proximal extragradient method, accelerated gradient method, inexact proximal method, saddle point problem, Nash equilibrium problem, Bregman distances.

LISTA DE TABELAS

TABELA 1 – Resultados dos testes para o problema bilinear	99
TABELA 2 – Resultados dos testes para o problema de ponto-de-sela vetor-matriz	101
TABELA 3 – Resultados dos testes para o problema quadrático	102

LISTA DE ABREVIATURAS E SIGLAS

HPE	Método híbrido proximal extragradiente
NE-HPE	Método híbrido proximal extragradiente não-Euclidiano
SP-HPE	Algoritmo HPE para resolver o problema de ponto-de-sela
ACC-HPE	Algoritmo HPE acelerado

LISTA DE SÍMBOLOS

∂f	Subdiferencial (usual) da função f .
$cl\ Z$	Fecho do conjunto Z .
$ri\ Z$	Interior relativo do conjunto Z .
$Gr\ T$	Gráfico do operador T (os pontos (x, y) onde $x \in T(y)$.)
\mathcal{I}_Z	Função indicadora do conjunto Z .
$f(\cdot, y)$	Dada uma função de duas variáveis $f : \mathcal{Z} \times \mathcal{Z} \mapsto \mathbb{R}$, fixado $y \in \mathcal{Z}$, $f(\cdot, y)$ denota a função de uma variável $\phi : \mathcal{Z} \mapsto \mathbb{R}$ dada por $\phi(x) =$ $f(x, y) \quad \forall x \in \mathcal{Z}$.
$argmin_{x \in X} \{f(x)\}$	Denota o ponto sobre o qual a função f atinge o valor mínimo sobre o conjunto X .
$\min_{x \in X} \{f(x)\}$	Denota o valor mínimo de f sobre o conjunto X .
$f \leq g$	Quando f e g são funções, denota que em cada ponto x na interseção dos domínios das funções vale $f(x) \leq g(x)$.
$\ \cdot\ _{\mathcal{X}}^*$	Denota a norma dual $\ \cdot\ _{\mathcal{X}}$ definida como
$\ x\ _{\mathcal{X}}^* := \max_{\ x'\ _{\mathcal{X}}=1} \{\langle x, x' \rangle_{\mathcal{X}} : x' \in \mathcal{X}\}, \quad \forall x \in \mathcal{X}.$	
$g(x) = \mathcal{O}(f(x))$	Quando existe uma constante positiva M tal que para todo valor suficientemente grande de x tem-se $ g(x) \leq Mf(x)$.
$[k]$	O menor número inteiro que não é menor que k .

SUMÁRIO

Lista de tabelas	vii
1 INTRODUÇÃO E RESULTADOS PRELIMINARES	1
1.1 Estrutura do Trabalho	1
1.2 Introdução	3
1.3 Funções Convexas e Operadores Monótonos	7
1.3.1 O Problema de Ponto-de-Sela	11
1.4 Algoritmo Acelerado Para Otimização Convexa Composta	18
1.5 Método HPE Para O Problema De Inclusão Monótona	26
1.6 Trabalhos Relacionados	27
2 UMA VERSÃO ACELERADA DO ALGORITMO HPE PARA O PROBLEMA DE PONTO-DE-SELA	30
2.1 O Algoritmo SP-HPE Para o Problema de Ponto-de-Sela	31
2.2 Uma Versão Acelerada do Método HPE Para o Problema de Ponto-de-Sela . . .	33
2.2.1 Resolvendo o Subproblema de HPE	34
2.3 Complexidade do Algoritmo Acelerado Para o Problema de Ponto-de-Sela . . .	40
2.4 Convergência do Algoritmo Acelerado Para o Problema de Ponto-de-Sela . . .	42
3 ALGORITMO HPE PARA O PROBLEMA DE EQUILÍBRIO DE NASH	48
3.1 O Problema de Equilíbrio de Nash	48
3.2 Uma Versão do Método HPE Para Resolver o Problema de Equilíbrio de Nash .	52

3.3 Um Algoritmo Acelerado Para o Problema de Equilíbrio de Nash	54
3.4 Resolvendo o Subproblema de HPE	59
3.4.1 Convergência do Algoritmo Acelerado Para o Problema de Equilíbrio de Nash	59
3.4.2 Um Segundo Algoritmo Acelerado Para Resolver o Problema de Equilíbrio de Nash	62
3.5 Complexidade dos Algoritmos Acelerados Para o Problema de Equilíbrio de Nash	69
4 ALGORITMO NÃO-EUCLIDIANO HÍBRIDO EXTRAGRADIEN-	
TE PROXIMAL	71
4.1 Método Não-Euclidiano Híbrido Extragradiante Proximal Para Problemas de Inclusão Monótona	71
4.2 Uma Condição Suficiente Para as Condições de Erro	78
5 UMA VERSÃO ACELERADA DO MÉTODO NE-HPE	84
5.1 Um Esquema Acelerado Para Resolver (P2)	85
5.2 Uma Versão Acelerada NE-HPE Para Resolver o Problema de Ponto-de-Sela . .	92
6 EXPERIMENTOS NUMÉRICOS	97
6.1 Problema Bilinear	99
6.2 Problema de Ponto-de-Sela Vetor-Matriz	100
6.3 Problema Quadrático	102
7 CONCLUSÃO E TRABALHOS FUTUROS	104
REFERÊNCIAS	106

1 INTRODUÇÃO E RESULTADOS PRELIMINARES

Neste capítulo introdutório especificamos a estrutura geral do trabalho, introduzimos notações e resultados de conhecimento geral e descrevemos dois algoritmos conhecidos da literatura: o algoritmo acelerado de Nesterov ([NESTEROV, 2005](#)) para otimização convexa e o método híbrido extragradiente proximal para o problema de inclusão monótona. O modo como tais algoritmos se relacionam com a proposta do trabalho será esclarecido mais adiante.

1.1 ESTRUTURA DO TRABALHO

A estrutura do trabalho é a seguinte: na Seção 1.2 descrevemos de modo geral a proposta do trabalho, na Seção 1.3 introduzimos notações, expondo os resultados básicos e propriedades de operadores ponto-conjunto e funções convexas e discutimos o problema de ponto-de-sela e algumas de suas propriedades, na Seção 1.4 revisamos um algoritmo de gradiente acelerado para resolver problemas de otimização convexa composta ([NESTEROV, 2005](#); [TSENG, 2008](#)), na Seção 1.5 elaboramos o método híbrido extragradiente proximal ([SOLODOV; SVAITER, 1999a](#)) para o problema de inclusão monótona e na Seção 1.6 apresentamos uma descrição geral dos trabalhos recentes relacionados com os problemas de interesse. No Capítulo 2 visamos a apresentar a generalização do algoritmo exibido em ([HE; MONTEIRO, 2014](#)) para uma classe mais geral de funções. Apresentamos o método de uma forma geral na Seção 2.1, na qual cada iteração possui um subproblema que deve ser resolvido com um algoritmo específico. Assim como em ([HE; MONTEIRO, 2014](#)), resolvemos tal subproblema fazendo uso de um algoritmo acelerado, a saber, aquele apresentado na Seção 1.4, e assim na Seção 2.2 apresentamos uma versão especial do método

na qual resolvemos o subproblema usando tal algoritmo. Os resultados de complexidade são enunciados na Seção 2.3, e deixamos a prova do resultado principal para a Seção 2.4. No Capítulo 3 apresentamos uma generalização do algoritmo em (HE; MONTEIRO, 2014) para o problema de equilíbrio de Nash de duas variáveis. na Seção 3.1 apresentamos o problema de equilíbrio de Nash. Na Seção 3.2 apresentamos o método HPE para este problema, e na Seção 3.3 derivamos uma versão acelerada de tal método, assim como feito no Capítulo 2. Na Seção 3.4 damos a prova de convergência da versão acelerada apresentada na Seção 3.3 e apresentamos uma outra variante acelerada para resolver o problema de equilíbrio de Nash que faz uso de uma minimização a menos, se comparada com a variante da Seção 3.3. Por fim, na Seção 3.5 discutimos resultados de complexidade das duas variantes. Nos dois capítulos seguintes visamos a apresentar a generalização do algoritmo de (HE; MONTEIRO, 2014) para o caso de distâncias de Bregman, conforme discutido anteriormente. O Capítulo 4 contém duas seções. Na Seção 4.1 apresentamos a noção de distâncias de Bregman, o método NE-HPE (SOLODOV; SVAITER, 2000) para resolver (1.3) e estabelecemos sua propriedade de complexidade ergódica. Na Seção 4.2 apresentamos uma condição suficiente para se obter a condição de erro (1.6), para uma classe particular de operadores monótonos (que inclui o operador dado por (1.1)). Já no Capítulo 5 descrevemos o novo algoritmo acelerado NE-HPE para resolver o problema de ponto-de-sela, i.e., a inclusão (1.3) com o operador dado por (1.1). Na Seção 5.1 apresentamos um algoritmo baseado no algoritmo de gradiente acelerado, da Seção 1.4, para determinar soluções aproximadas do sub-problema de inclusão relacionado à condição de erro do método NE-HPE em (1.6), e derivamos sua complexidade. Na Seção 5.2 descrevemos o novo método NE-HPE acelerado para resolver o problema de ponto-de-sela e estabelecemos sua complexidade. Nesta seção também discutimos uma maneira de escolher o tamanho do passo λ , de modo que a complexidade interna se reduza à obtida para o algoritmo de suavização de Nesterov

em (NESTEROV, 2005). Finalmente, apresentamos resultados numéricos no Capítulo 6, mostrando que o algoritmo proposto para o problema de ponto-de-sela teve desempenho melhor que o algoritmo de suavização de Nesterov de (NESTEROV, 2005) em três classes de problemas de ponto-de-sela e finalizamos o trabalho com o Capítulo 7, apresentando conclusões sobre o trabalho e perspectivas de desenvolvimentos futuros.

1.2 INTRODUÇÃO

Nosso objetivo neste trabalho é propor algoritmos para resolver os problemas de ponto-de-sela e de equilíbrio de Nash. Dados conjuntos convexos não-vazios $X \subset \mathcal{X}$ e $Y \subset \mathcal{Y}$ onde \mathcal{X} e \mathcal{Y} são dois espaços com produto interno, e uma função $\hat{\Phi} : X \times Y \rightarrow \mathbb{R}$, iremos propor um algoritmo para determinar, quando existirem, pontos de sela para $\hat{\Phi}$, i.e., pares $(\bar{x}, \bar{y}) \in X \times Y$ tais que

$$\hat{\Phi}(\bar{x}, y) \leq \hat{\Phi}(\bar{x}, \bar{y}) \leq \hat{\Phi}(x, \bar{y}), \quad \forall (x, y) \in X \times Y,$$

ou equivalentemente, um zero do operador $T : \mathcal{X} \times \mathcal{Y} \rightrightarrows \mathcal{X} \times \mathcal{Y}$ definido como

$$T(x, y) = \begin{cases} \partial[\hat{\Phi}(\cdot, y) - \hat{\Phi}(x, \cdot)](x, y), & \forall (x, y) \in X \times Y; \\ \emptyset, & \text{caso contrário.} \end{cases} \quad (1.1)$$

Veremos que com algumas hipóteses sobre $\hat{\Phi}$, o operador T é monótono maximal, e portanto um zero aproximado de T pode ser computado com o uso de um método proximal inexato tal como um dos algoritmos apresentados em (HE; MONTEIRO, 2014; KORPELEVIČ, 1976; MONTEIRO; SVAITER, 2011; MONTEIRO; SVAITER, 2010b; MONTEIRO; SVAITER, 2013; NEMIROVSKI, 2004; SOLODOV; SVAITER, 1999a; SOLODOV; SVAITER, 1999b; SOLODOV; SVAITER, 2000; SOLODOV; SVAITER, 2001; TSENG, 2000).

Em particular, He e Monteiro (HE; MONTEIRO, 2014) propuseram um método proximal inexato para resolver um caso especial do problema de ponto-de-sela, quando $\hat{\Phi}$

é da forma bilinear

$$\hat{\Phi}(x, y) = f(x) + \langle Ax, y \rangle + g_1(x) - g_2(y); \quad (1.2)$$

e $A : \mathcal{X} \rightarrow \mathcal{Y}$ é um operador linear, g_1 e g_2 são funções convexas, próprias e fechadas, f é uma função convexa e diferenciável cujo gradiente é Lipschitziano e $X \times Y = \text{dom } g_1 \times \text{dom } g_2$. O método é um caso especial do método híbrido extragradiente proximal (HPE) introduzido em (SOLODOV; SVAITER, 1999a). Qualquer versão do método HPE é essencialmente um método proximal inexato que permite um erro nos sub-problemas de inclusão. Mais especificamente, dado um operador monótono maximal T , considere o problema de inclusão monótona

$$0 \in T(z). \quad (1.3)$$

A cada iteração do método proximal clássico, dado um par (z_-, λ) , com $\lambda > 0$, é computado um z como a (única) solução de

$$\frac{z_- - z}{\lambda} \in T(z), \quad (1.4)$$

ou equivalentemente, quando o operador T for inversível, $z = (\lambda T + I)^{-1}(z_-)$. Por outro lado, uma versão do método HPE, a cada iteração determina uma solução aproximada de (1.3) satisfazendo um critério de erro (relativo), a saber, para uma tolerância $\sigma \in [0, 1]$, a tripla $(\tilde{z}, z, \varepsilon)$ é computada de modo a satisfazer

$$\frac{z_- - z}{\lambda} \in T^\varepsilon(\tilde{z}) \quad \text{e} \quad \frac{1}{2}\|\tilde{z} - z\|^2 + \lambda\varepsilon \leq \frac{1}{2}\sigma^2\|\tilde{z} - z_-\|^2, \quad (1.5)$$

onde T^ε denota o ε -enlargamento (BURACHIK; IUSEM; SVAITER, 1997) de T . (Como veremos na Seção 1.3, ele tem a propriedade de que $T^\varepsilon(u) \supset T(u)$ para todo u , e vale a igualdade quando $\varepsilon = 0$). Fazendo $\sigma = 0$ em (1.5), temos que $z = \tilde{z}$ e $\varepsilon = 0$, assim a inclusão em (1.5) se reduz a (1.4). Ao contrário de outros métodos HPE na literatura

(como por exemplo (HE; MONTEIRO, 2013; MONTEIRO; SVAITER, 2013)) em que temos que escolher λ relativamente pequeno, no método HPE de (HE; MONTEIRO, 2014) para resolver (1.3) com T dado por (1.1), podemos escolher um $\lambda > 0$ arbitrário e computar a tripla $(\tilde{z}, z, \varepsilon)$ satisfazendo a condição (1.5) usando um algoritmo de gradiente acelerado (e.g., (NESTEROV, 2005; TSENG, 2008)), aplicado a um certo problema de máximo-mínimo regularizado associado a $\hat{\Phi}$ dada por (1.2).

Nossas contribuições com respeito ao problema de ponto-de-sela, neste trabalho, consistem em desenvolver generalizações do algoritmo proposto em (HE; MONTEIRO, 2014) de duas maneiras. Primeiro, construímos uma generalização resolvendo problemas de ponto-de-sela com quaisquer funções $\hat{\Phi}$ cujo gradiente $\nabla_x \hat{\Phi}$ é Lipschitziano, ao invés de usar as funções $\hat{\Phi}$ dadas por (1.2). Segundo, com apenas algumas hipóteses adicionais, apresentamos outra generalização que, além de fazer uso de funções quaisquer, usa distâncias de Bregman gerais ao invés da Euclidiana para especificar o erro relativo do sub-problema da iteração. Mais especificamente, consideramos $dw_z(z') = w(z') - w(z) - \langle \nabla w(z), z' - z \rangle$ para cada z, z' , em que w é uma função convexa diferenciável. Então, a distância Euclidiana é obtida fazendo $w(\cdot) = \|\cdot\|^2/2$ e $dw_z(z') = \|z' - z\|^2/2$ para todo z', z e podemos reescrever (1.5) como

$$\frac{1}{\lambda} \nabla(dw)_z(z_-) \in T^\varepsilon(\tilde{z}), \quad (dw)_z(\tilde{z}) + \lambda\varepsilon \leq \sigma(dw)_{z_-}(\tilde{z}). \quad (1.6)$$

O método HPE não-Euclidiano (NE-HPE) (SOLODOV; SVAITER, 2000) generaliza o HPE no sentido de que determina uma solução aproximada de (1.3) satisfazendo a condição mais geral (1.6), com w sendo uma função diferenciável convexa arbitrária. Um passo importante em analisar o método NE-HPE consiste em estabelecer uma propriedade de complexidade ergódica para o método NE-HPE para resolver (1.3), quando T é um operador monótono maximal com domínio limitado. Similarmente ao método em (HE;

(MONTEIRO, 2014), no método NE-HPE podemos escolher $\lambda > 0$ e computar uma tripla $(\tilde{z}, z, \varepsilon)$ satisfazendo a condição de erro (1.6), usando um algoritmo de gradiente acelerado aplicado a um problema associado de máximo-mínimo regularizado. Supondo que o conjunto viável do problema de ponto-de-sela é limitado, um limite superior de complexidade ergódica será determinado para o número total de iterações internas (do algoritmo de gradiente acelerado) do algoritmo proposto. Finalmente, mostraremos que se o tamanho do passo λ e a distância de Bregman forem escolhidas convenientemente, então a complexidade ergódica determinada reduz-se à obtida em (NESTEROV, 2005) para o algoritmo de suavização de Nesterov, que determina soluções aproximadas para problemas de ponto-de-sela com estrutura bilinear. Tal complexidade é ótima, conforme veremos na Seção 5.2.

Outra contribuição deste trabalho é aplicar a mesma ideia utilizada para se resolver o problema de ponto-de-sela para problemas de equilíbrio de Nash.

Dados $X \subseteq \mathcal{X}$ e $Y \subseteq \mathcal{Y}$ não-vazios, convexos e fechados, considere duas funções $\hat{\Phi}_i : X \times Y \rightarrow \mathbb{R}$, $i = 1, 2$. O problema de equilíbrio de Nash de duas variáveis determinado por $(\hat{\Phi}_1, \hat{\Phi}_2)$ e $Z = X \times Y$ consiste em determinar um par $(x, y) \in X \times Y$ tal que

$$\hat{\Phi}_1(x, y) \leq \hat{\Phi}_1(x', y), \quad \hat{\Phi}_2(x, y) \leq \hat{\Phi}_2(x, y'), \quad \forall (x', y') \in X \times Y.$$

Assim como no caso do problema de ponto-de-sela, temos que se $(x, y) \in X \times Y$ e

$$(0, 0) \in \partial[\hat{\Phi}_1(\cdot, y) + \hat{\Phi}_2(x, \cdot)](x, y), \tag{1.7}$$

então (x, y) é um ponto de equilíbrio de Nash, ou seja, o problema também pode ser reduzido a um problema de inclusão monótona. A ideia utilizada para se obter um algoritmo para encontrar uma solução aproximada para este problema é a mesma utilizada no contexto do problema de ponto-de-sela, isto é, reduzir o problema ao problema de

inclusão monótona determinado pelo operador (1.7) e então combinar o algoritmo HPE com um método acelerado para resolver o subproblema interno.

1.3 FUNÇÕES CONVEXAS E OPERADORES MONÓTONOS

Nesta seção apresentamos notações e definições que serão usadas durante o trabalho. Os resultados mencionados nestas seções são conhecidos na literatura especializada. Para maiores detalhes consulte ([AUSLENDER; TEBoulLE, 2003](#)).

Denotamos por \mathbb{R} o conjunto dos números reais e denotamos o conjunto dos números reais não negativos e o dos reais positivos respectivamente como \mathbb{R}_+ e \mathbb{R}_{++} . $\lceil z \rceil$ denota o menor inteiro que não é menor que $z \in \mathbb{R}$.

Considere \mathcal{Z} um espaço de dimensão finita com um produto interno denotado por $\langle \cdot, \cdot \rangle$. Dado um conjunto $Z \subset \mathcal{Z}$, denotamos seu interior relativo por $\text{ri } Z$ e seu fecho por $\text{cl } Z$. Podemos identificar uma relação $T \subseteq \mathcal{Z} \times \mathcal{Z}$ com um operador ponto-conjunto $T : \mathcal{Z} \rightrightarrows \mathcal{Z}$ tal que

$$T(z) := \{v \in \mathcal{Z} : (z, v) \in T\}, \quad \forall z \in \mathcal{Z}.$$

O gráfico do operador $T : \mathcal{Z} \rightrightarrows \mathcal{Z}$ é dado por

$$\text{Gr}(T) := \{(z, v) \in \mathcal{Z} \times \mathcal{Z} : v \in T(z)\}.$$

Definimos o domínio de T como

$$\text{dom } T := \{z \in \mathcal{Z} : T(z) \neq \emptyset\}.$$

Denotamos o domínio de um operador ponto a ponto (função) F também por $\text{dom } F$. Um operador $T : \mathcal{Z} \rightrightarrows \mathcal{Z}$ é monótono se

$$\langle v - \tilde{v}, z - \tilde{z} \rangle \geq 0, \quad \forall (z, v), (\tilde{z}, \tilde{v}) \in \text{Gr}(T).$$

Além disso, T é monótono maximal se é monótono e maximal na família de operadores monótonos com respeito à ordem parcial de inclusão, i.e., $S : \mathcal{Z} \rightrightarrows \mathcal{Z}$ monótono e $\text{Gr}(S) \supset \text{Gr}(T)$ implica que $S = T$.

O seguinte resultado estabelece a convexidade do fecho do domínio de um operador monótono maximal

Proposição 1.1 *Considere um operador monótono maximal $T : \mathcal{Z} \rightrightarrows \mathcal{Z}$, então $\text{cl dom } T$ é convexo.*

Prova. Veja Teorema 1 de (ROCKAFELLAR, 1970d). ■

Dado um escalar $\varepsilon \geq 0$, o ε -enlargamento de um operador ponto-conjunto (BURACHIK; IUSEM; SVAITER, 1997) $T : \mathcal{Z} \rightrightarrows \mathcal{Z}$ é o operador ponto-conjunto $T^\varepsilon : \mathcal{Z} \rightrightarrows \mathcal{Z}$ dado por

$$T^\varepsilon(z) := \{v \in \mathcal{Z} : \langle z - \tilde{z}, v - \tilde{v} \rangle \geq -\varepsilon, \quad \forall \tilde{z} \in \mathcal{Z}, \forall \tilde{v} \in T(\tilde{z})\} \quad \forall z \in \mathcal{Z}. \quad (1.8)$$

Pela definição acima, vemos que para todo $\varepsilon \geq 0$ vale $T^\varepsilon \supset T$ para todo operador ponto-conjunto T . Para operadores monótonos maximais, temos ainda que $T^0 = T$. Com efeito, $T^0 \supset T$ pela observação anterior. Tomando $v \in T^0(z)$, temos que

$$\langle z - z', v - v' \rangle \geq 0, \quad \forall (z', v') \in \text{Gr}(T) \subset \text{Gr}(T^0).$$

Pela arbitrariedade de (z, v) segue que T^0 é monótono, e pela maximalidade de T , devemos ter que $T^0 = T$.

O seguinte resultado estabelece propriedades mais gerais de ε -enlargamentos. A prova também segue imediatamente da definição (1.8).

Proposição 1.2 *Considere $T, T' : \mathcal{Z} \rightrightarrows \mathcal{Z}$ operadores. Então, valem as seguintes propriedades:*

(a) $T^{\varepsilon_1}(z) + (T')^{\varepsilon_2}(z) \subset (T + T')^{\varepsilon_1 + \varepsilon_2}(z)$ para todo $z \in \mathcal{Z}$ e $\varepsilon_1, \varepsilon_2 \in \mathbb{R}$.

(b) Se, além disso, T_1 e T_2 são operadores monótonos e $\lambda, \mu \geq 0$, então $\lambda T_1 + \mu T_2$ também é um operador monótono.

Prova. A prova de (a) segue diretamente de (1.8). A prova de (b) pode ser encontrada na Proposição 6.1.1 (b) de (AUSLENDER; TEBoulLE, 2003). ■

Considere uma função $f : \mathcal{Z} \rightarrow [-\infty, \infty]$. Definimos o domínio efetivo de f como

$$\text{dom } f := \{z \in \mathcal{Z} : f(z) < \infty\}.$$

Dado um escalar $\varepsilon \geq 0$, definimos o ε -subdiferencial de f denotado por $\partial_\varepsilon f : \mathcal{Z} \rightrightarrows \mathcal{Z}$ como

$$\partial_\varepsilon f(z) := \{v : f(\tilde{z}) \geq f(z) + \langle \tilde{z} - z, v \rangle - \varepsilon, \forall \tilde{z} \in \mathcal{Z}\} \quad \forall z \in \mathcal{Z}. \quad (1.9)$$

Quando $\varepsilon = 0$, denotamos o operador $\partial_\varepsilon f$ simplesmente como ∂f e nos referimos a ele como o subdiferencial de f .

O seguinte resultado foi enunciado e provado com um erro em (ROCKAFELLAR, 1966), e posteriormente corrigido em (ROCKAFELLAR, 1970b).

Proposição 1.3 *O operador ∂f é monótono se f é própria e é monótono maximal se f é própria, convexa e fechada.*

Na próxima proposição enunciamos a relação entre o subdiferencial de uma função dada pela soma de duas funções convexas e a soma dos subdiferenciais de cada uma.

Proposição 1.4 *Considere f_1, f_2 funções convexas próprias definidas em \mathcal{Z} e $f = f_1 + f_2$.*

Então, temos que

$$(a) \quad \partial f(x) \supset \partial f_1(x) + \partial f_2(x), \quad \forall x \in \mathcal{Z}.$$

Se $\text{ri dom } f_1 \cap \text{ri dom } f_2 \neq \emptyset$ então temos ainda que

$$(b) \quad \partial f(x) = \partial f_1(x) + \partial f_2(x), \quad \forall x \in \mathcal{Z}.$$

Prova. (a) Dado $x \in \mathcal{Z}$, tome $x_i \in \partial f_i(x), i = 1, 2$. Então para todo $u \in \mathcal{Z}$

$$\begin{aligned} f(u) &= f_1(u) + f_2(u) \geq f_1(x) + \langle u - x, x_1 \rangle + f_2(x) + \langle u - x, x_2 \rangle \\ &= f(x) + \langle u - x, x_1 + x_2 \rangle, \end{aligned}$$

donde $x_1 + x_2 \in \partial f(x)$.

(b) Ver o Teorema 23.8 de (ROCKAFELLAR, 1970a). ■

Podemos generalizar a Proposição 1.4 para operadores monótonos quaisquer ao invés de subdiferenciais de funções convexas, como percebemos no próximo resultado.

Proposição 1.5 *Considere T_1 e $T_2 : \mathcal{Z} \rightrightarrows \mathcal{Z}$ dois operadores monótonos maximais. Se $\text{ri dom } T_1 \cap \text{ri dom } T_2 \neq \emptyset$ então $T_1 + T_2$ também é monótono maximal.*

Prova. Ver o Teorema 2 de (ROCKAFELLAR, 1970c). ■

Uma propriedade importante do ε -subdiferencial é a sua relação com um operador monótono dado pelo subdiferencial de uma função convexa. Neste caso, o ε -subdiferencial está contido no ε -enlargamento do subdiferencial.

Proposição 1.6 *Considere f uma função convexa e $T = \partial f$, então $\partial_\varepsilon f(x) \subset (\partial f)^\varepsilon(x), \quad \forall x \in \mathcal{Z}, \varepsilon \geq 0$.*

Prova. Tome $u \in \partial_\varepsilon f(x)$ e considere $(y, v) \in \text{Gr } T$, com $T = \partial f$. Pela definição de $\partial_\varepsilon f$ e ∂f , temos que

$$\varepsilon + \langle u, x - y \rangle \geq f(x) - f(y) \geq \langle v, x - y \rangle,$$

o que implica que $\langle u - v, x - y \rangle \geq -\varepsilon$, donde $u \in T^\varepsilon(x)$. ■

Dado um conjunto $\Omega \subset \mathcal{Z}$, definimos a função indicadora $\mathcal{I}_\Omega : \mathcal{Z} \rightarrow (-\infty, \infty]$ de Ω como

$$\mathcal{I}_\Omega(z) := \begin{cases} 0, & z \in \Omega, \\ \infty, & z \notin \Omega. \end{cases} \quad (1.10)$$

O resultado seguinte mostra que adicionando um operador monótono maximal T ao subdiferencial de uma função indicadora de um conjunto convexo contendo $\text{dom } T$, não se modifica T .

Proposição 1.7 *Considere T um operador monótono maximal e $\Omega \subset \mathcal{Z}$ um subconjunto convexo contendo $\text{dom } T$. Então, $T + \partial\mathcal{I}_\Omega = T$.*

Prova. Temos que $\partial\mathcal{I}_\Omega$ é monótono, pois por hipótese, o conjunto Ω é convexo, e portanto a função indicadora \mathcal{I}_Ω é convexa. Uma vez que T também é monótono por hipótese, segue que $T + \partial\mathcal{I}_\Omega$ é monótono (pela Proposição 1.2 (b)). Como $\text{dom } T \subset \Omega$ e $0 \in \partial\mathcal{I}_\Omega(x)$ para todo $x \in \Omega$, temos que $T \subset T + \partial\mathcal{I}_\Omega$. A inclusão $T \supset T + \partial\mathcal{I}_\Omega$ segue da hipótese de que T é monótono maximal. Sendo assim temos que $T = T + \partial\mathcal{I}_\Omega$. ■

Teorema 1.1 *Considere um operador monótono $T : \mathcal{Z} \rightrightarrows \mathcal{Z}$ e $\lambda > 0$. Então T é maximal se e somente se $\text{dom}(I + \lambda T)^{-1} = \mathcal{Z}$.*

Prova. Este resultado é conhecido como o Teorema de Minty, cuja prova pode ser encontrada em (MINTY, 1978). ■

1.3.1 O Problema de Ponto-de-Sela

Considere \mathcal{X} e \mathcal{Y} espaços com produto interno de dimensão finita cujos produtos internos denotamos respectivamente por $\langle \cdot, \cdot \rangle_{\mathcal{X}}$ e $\langle \cdot, \cdot \rangle_{\mathcal{Y}}$. Induzimos o espaço $\mathcal{X} \times \mathcal{Y}$ com o

produto interno canônico definido como

$$\langle (x, y), (x', y') \rangle = \langle x, x' \rangle_{\mathcal{X}} + \langle y, y' \rangle_{\mathcal{Y}}, \quad \forall (x, y), (x', y') \in \mathcal{X} \times \mathcal{Y}.$$

Considere $X \subseteq \mathcal{X}$ e $Y \subseteq \mathcal{Y}$ conjuntos não-vazios e considere uma função $\hat{\Phi} : \text{dom } \hat{\Phi} \rightarrow \mathbb{R}$ tal que $\text{dom } \hat{\Phi} \supset X \times Y$.

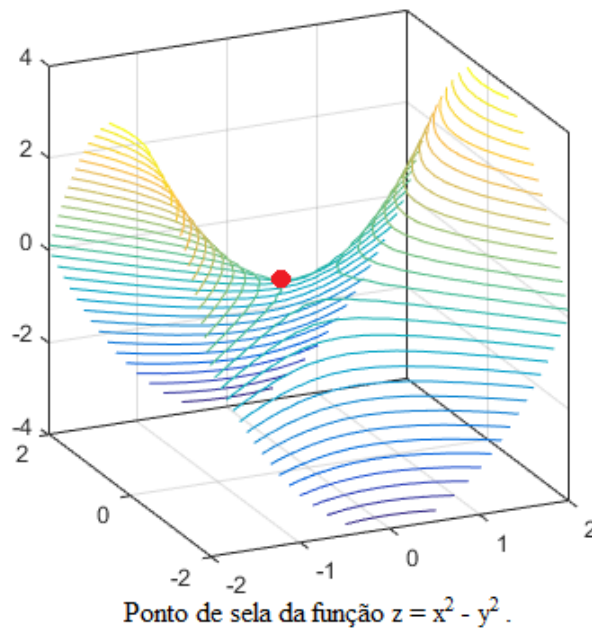
O par $(\bar{x}, \bar{y}) \in X \times Y$ é um ponto-de-sela de $\hat{\Phi}$ com respeito a $X \times Y$ se satisfaz as desigualdades

$$\hat{\Phi}(\bar{x}, y) \leq \hat{\Phi}(\bar{x}, \bar{y}) \leq \hat{\Phi}(x, \bar{y}), \quad \forall (x, y) \in X \times Y.$$

O problema de determinar tal par será chamado de problema de ponto-de-sela de $\hat{\Phi}$ com respeito à $Z := X \times Y$ e denotado por $\text{SP}(\hat{\Phi}; Z)$.

A superfície de um ponto-de-sela tem este nome por lembrar uma sela, côncava em uma direção e convexa em uma ou mais direções, conforme mostra a figura abaixo.

FIGURA 1 – Figura: O Ponto de Sela



A ideia de ponto-de-sela tem várias aplicações. Uma delas é o uso do problema de maximização de utilidade do consumidor, da economia. Suponha que um consumidor esteja interessado em dois tipos de produtos denotados por p e q com quantidades denotadas por

P e Q , respectivamente, com função de utilidade $U(p, q)$ expressando sua preferência pelos produtos. Naturalmente, com uma quantia que esteja disposto a pagar dada por $M \geq 0$, o consumidor deseja maximizar o ganho de produtos p e q , em outras palavras, o problema de maximização de utilidade do consumidor é dado, neste caso, por:

$$\max U(p, q), \quad \text{s. a } Pp + Qq \leq M.$$

Aplicando as ideias da otimização, podemos resolver tal problema usando a função lagrangeana

$$L(p, q, \lambda) = U(p, q) - \lambda(Pp + Qq - M).$$

Derivando L com respeito a cada variável, conseguimos as condições necessárias de primeira ordem de Karush-Kuhn-Tucker

$$\frac{\partial L}{\partial p} = \frac{\partial U}{\partial p} - \lambda P = 0, \quad (1.11)$$

$$\frac{\partial L}{\partial q} = \frac{\partial U}{\partial q} - \lambda Q = 0, \quad (1.12)$$

$$\frac{\partial L}{\partial \lambda} = -(Pp + Qq - M) \geq 0$$

$$\lambda(Pp + Qq - M) = 0, \quad \lambda \geq 0 \quad (1.13)$$

Se $\lambda = 0$, temos $-(Pp + Qq - M) \geq 0$, donde o problema é irrestrito, portanto pode-se resolver o problema apenas igualando-se as duas derivadas de U a zero. Por outro lado, se $\lambda > 0$, temos $-(Pp + Qq - M) = 0$, o problema pode ser resolvido pelo sistema dado por (1.11), (1.12) e (1.13).

A solução ótima (p^*, q^*, λ^*) deste problema está no ponto-de-sela da lagrangeana. Neste caso a superfície de L se curva para cima nas direções dos produtos p e q e se curva para baixo na direção do multiplicador λ . Desta forma o problema pode ser escrito como um SP($\hat{\Phi}; Z$) com $\hat{\Phi} = L$ e $Z = \mathbb{R}^3$.

O problema de ponto-de-sela $\text{SP}(\hat{\Phi}; Z)$ pode ser escrito na forma de um problema de inclusão monótona. Tomando $\mathcal{Z} := \mathcal{X} \times \mathcal{Y}$ e definindo para todo $z = (x, y) \in Z$, a função

$$\hat{\Phi}_z(x', y') = \begin{cases} \hat{\Phi}(x', y) - \hat{\Phi}(x, y'), & \text{se } (x', y') \in Z; \\ +\infty, & \text{caso contrário.} \end{cases} \quad (1.14)$$

e um operador $T_{\hat{\Phi}} : \mathcal{Z} \rightrightarrows \mathcal{Z}$ como

$$T_{\hat{\Phi}}(z) := \begin{cases} \partial(\hat{\Phi}_z)(z), & \text{se } z \in Z; \\ \emptyset, & \text{caso contrário,} \end{cases} \quad (1.15)$$

podemos ver que se $z = (x, y)$ é uma solução da inclusão

$$0 \in T_{\hat{\Phi}}(z), \quad (1.16)$$

então z é um ponto-de-sela de $\hat{\Phi}$ com respeito a Z . De fato, se $z = (x, y)$ é solução de (1.16) então para todo $z' = (x', y') \in Z$ temos que

$$0 \in \partial [\hat{\Phi}(x', y) - \hat{\Phi}(x, y')],$$

o que implica, pela relação (1.9), que

$$\hat{\Phi}(x', y) - \hat{\Phi}(x, y') \geq 0 + \langle 0, (x', y') - (x, y) \rangle = 0,$$

e como esta relação vale para todo (x', y') , segue que

$$\hat{\Phi}(x, y') \leq \hat{\Phi}(x, y) \leq \hat{\Phi}(x', y).$$

Em vista desta observação, a seguinte definição descreve a noção de solução aproximada de $\text{SP}(\hat{\Phi}; Z)$ considerada em nossa análise.

Definição 1.1 *Dado $\bar{\varepsilon}, \rho \geq 0$, chamamos a tripla $(z, v, \varepsilon) \in Z \times Z \times \mathbb{R}_+$ satisfazendo $v \in \partial_{\varepsilon}(\hat{\Phi}_z)(z)$ com $\|v\| \leq \rho$ de $(\rho, \bar{\varepsilon})$ -ponto-de-sela de $\hat{\Phi}$ com respeito a Z se $\varepsilon \leq \bar{\varepsilon}$, sendo $\hat{\Phi}_z$ dada por (1.14). Se $v = 0$, dizemos apenas que (z, ε) é um $\bar{\varepsilon}$ -ponto-de-sela de $\hat{\Phi}$ com respeito à Z .*

O operador $T_{\hat{\Phi}}$ admite o ε -enlargamento como em (1.8). Ele também admite o ε -subdiferencial $\partial_\varepsilon(\hat{\Phi}_z)(z)$ para $z \in Z$ que explora sua natureza de ponto-de-sela. O seguinte resultado segue diretamente das relações (1.8) e (1.9).

Proposição 1.8 *Para cada $z \in Z$ e $\varepsilon \geq 0$, vale a inclusão $\partial_\varepsilon(\hat{\Phi}_z)(z) \subset [T_{\hat{\Phi}}]^\varepsilon(z)$.*

Prova. Por argumento de vacuidade, basta considerar o caso em que $z = (x, y)$ está no conjunto Z . Considere $z = (x, y) \in Z$ e $v \in \partial_\varepsilon [\hat{\Phi}(\cdot, y) - \hat{\Phi}(x, \cdot)](x, y)$. Pela relação (1.9) isto implica que

$$\hat{\Phi}(x', y) - \hat{\Phi}(x, y') \geq \langle v, (x'y') - (x, y) \rangle - \varepsilon, \quad \forall (x', y') \in Z. \quad (1.17)$$

Por outro lado, pela definição de $T_{\hat{\Phi}}$, temos que para todo $v' \in T_{\hat{\Phi}}(x', y')$ vale

$$\hat{\Phi}(x, y') - \hat{\Phi}(x', y) \geq \langle v', (x, y) - (x', y') \rangle. \quad (1.18)$$

Somando (1.17) e (1.18), obtemos

$$\langle (x, y) - (x', y'), v - v' \rangle \geq -\varepsilon, \quad \forall (x', y') \in Z, v' \in T_{\hat{\Phi}}(x', y').$$

Segue, pela definição de $T_{\hat{\Phi}}$, que $v \in T^\varepsilon(x, y)$, conforme queríamos demonstrar. \blacksquare

O resultado seguinte dá condições suficientes para que o operador $T_{\hat{\Phi}}$ em (1.15) seja monótono maximal.

Proposição 1.9 *Valem os seguintes itens:*

(a) $T_{\hat{\Phi}}$ é monótono;

(b) se a função $\hat{\Phi}_z : \mathcal{Z} \rightarrow (-\infty, +\infty]$ é convexa e fechada para todo $z \in Z$, então $T_{\hat{\Phi}}$ é monótono maximal.

Prova.

a) Uma vez que o operador é vazio fora de Z , por um argumento de vacuidade basta mostrar que vale a desigualdade $\langle v_1 - v_2, z_1 - z_2 \rangle \geq 0$ para $z_i \in Z, v_i \in T(z_i), \quad i = 1, 2$. Tome $z_1 = (x_1, y_1), z_2 = (x_2, y_2) \in Z$. Suponha que

$$v_1 \in \partial [\hat{\Phi}(\cdot, y_1) - \hat{\Phi}(x_1, \cdot)](x_1, y_1),$$

$$v_2 \in \partial [\hat{\Phi}(\cdot, y_2) - \hat{\Phi}(x_2, \cdot)](x_2, y_2).$$

Pela relação (1.9) temos que para todo $(u, v) \in Z$ valem

$$\hat{\Phi}(u, y_1) - \hat{\Phi}(x_1, v) \geq \langle v_1, (u, v) - (x_1, y_1) \rangle,$$

$$\hat{\Phi}(u, y_2) - \hat{\Phi}(x_2, v) \geq \langle v_2, (u, v) - (x_2, y_2) \rangle.$$

Em particular, fazendo $(u, v) = (x_2, y_2)$ na primeira desigualdade e $(u, v) = (x_1, y_1)$ na segunda e somando, temos que

$$0 \geq \langle v_2 - v_1, z_1 - z_2 \rangle.$$

Segue que o operador $T_{\hat{\Phi}}$ é monótono.

b) Ver Teorema 6.3.2 de ([AUSLENDER; TEBoulle, 2003](#)), e também toda a discussão precedente. ■

Note que, pela definição de T^ε , a verificação da inclusão $v \in T^\varepsilon(x)$ requer a checagem de um grande número de desigualdades. Esta verificação só é viável para estruturas especiais de operadores T . Contudo, é possível computar pontos no gráfico de T^ε usando a seguinte fórmula de transporte fraco. A seguinte proposição foi obtida por Burachik et al. em ([BURACHIK; SAGASTIZÁBAL; SVAITER, 1999](#)), apresentamos a prova aqui por conveniência.

Proposição 1.10 *Suponha que $T : \mathcal{Z} \rightrightarrows \mathcal{Z}$ é monótono maximal. Para $i = 1, \dots, k$, tome*

$z_i, r_i \in \mathcal{Z}$ e $\varepsilon_i, \alpha_i \in \mathbb{R}_+$ satisfazendo $\sum_{i=1}^k \alpha_i = 1$ e defina

$$z^a = \sum_{i=1}^k \alpha_i z_i, \quad r^a = \sum_{i=1}^k \alpha_i r_i, \quad \varepsilon^a = \sum_{i=1}^k \alpha_i [\varepsilon_i + \langle z_i - z^a, r_i - r^a \rangle].$$

Então, valem os seguintes itens:

(a) *se $r_i \in T^{\varepsilon_i}(z_i)$ para todo $i = 1, \dots, k$, então $\varepsilon^a \geq 0$ e $r^a \in T^{\varepsilon^a}(z^a)$;*

(b) *se $T = T_{\hat{\Phi}}$ onde $\hat{\Phi}$ é uma função tal que para todo $z \in Z$ sua função estendida $\hat{\Phi}_z$ dada por (1.14) é convexa e fechada, e a inclusão $r_i \in \partial_{\varepsilon_i}(\hat{\Phi}_{z_i})(z_i)$ vale para cada $i = 1, \dots, k$, então*

$$r^a \in \partial_{\varepsilon^a}(\hat{\Phi}_{z^a})(z^a).$$

Prova. Provaremos o item (a). A prova de (b) é similar. Pela hipótese, definições de z^a , r^a e ε^a e pela relação (1.8) temos que para todo par $(z, r) \in \text{Gr}T$ vale

$$\begin{aligned} \langle z^a - z, r^a - r \rangle &= \sum_{i=1}^k \alpha_i \langle z_i - z, r^a - r \rangle \\ &= \sum_{i=1}^k \alpha_i [\langle z_i - z, r^a - r_i \rangle + \langle z_i - z, r_i - r \rangle] \\ &\geq \sum_{i=1}^k \alpha_i \langle z_i - z, r^a - r_i \rangle - \sum_{i=1}^k \alpha_i \varepsilon_i. \end{aligned} \tag{1.19}$$

Por outro lado, pela definição de r^a , temos que

$$\begin{aligned} \sum_{i=1}^k \alpha_i \langle z_i - z, r^a - r_i \rangle &= \sum_{i=1}^k \alpha_i [\langle z_i - z^a, r^a - r_i \rangle + \langle z^a - z, r^a - r_i \rangle] \\ &= - \sum_{i=1}^k \alpha_i \langle z_i - z^a, r_i - r^a \rangle. \end{aligned} \tag{1.20}$$

Combinando (1.19) e (1.20) e usando a definição de ε^a , temos que

$$\langle z^a - z, r^a - r \rangle \geq -\varepsilon^a,$$

e com isso vemos que vale $r^a \in T^{\varepsilon^a}(z^a)$.

Falta provar que $\varepsilon \geq 0$. Suponha por contradição que $\varepsilon < 0$, então temos que

$$\langle z^a - z, r^a - r \rangle > 0, \quad \forall (z, r) \in GrT.$$

Em particular, fazendo $(z, r) = (z^a, r^a)$, temos que $0 > 0$, absurdo, completando a prova.

■

1.4 ALGORITMO ACELERADO PARA OTIMIZAÇÃO CONVEXA COMPOSTA

Nesta subseção apresentamos uma variante do algoritmo acelerado de Nesterov ([NESTEROV, 2005](#); [TSENG, 2008](#)) para minimizar uma função (possivelmente fortemente) convexa e composta. Dizemos que uma função é fortemente convexa de constante $\mu > 0$ em um subconjunto Z de seu domínio, ou apenas uma função μ -fortemente convexa em Z se para todo $x, x' \in Z$ e $\alpha \in [0, 1]$ vale

$$f(\alpha x + (1 - \alpha)x') \leq \alpha f(x) + (1 - \alpha)f(x') - \frac{\mu\alpha(1 - \alpha)}{2} \|x - x'\|^2.$$

Se f for diferenciável, isto é equivalente a

$$f(x') \geq f(x) + \langle \nabla f(x), x' - x \rangle - \frac{\mu}{2} \|x' - x\|^2.$$

No mesmo contexto, dizemos que uma função f é μ -fortemente côncava se a função $-f$ é μ -fortemente convexa. Uma propriedade já conhecida de funções fortemente convexas é a seguinte:

Proposição 1.11 *Considere um conjunto convexo não-vazio $Z \subset \mathcal{Z}$. $f : \mathcal{Z} \mapsto \mathbb{R}$ é μ -fortemente convexa em Z se e somente se a função $f - \frac{\mu}{2} \|\cdot\|^2$ é convexa em Z .*

Prova. Por definição, f é μ -fortemente convexa se e somente se para todo $\alpha \in [0, 1]$ vale

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y) - \frac{\mu}{2} \alpha(1 - \alpha) \|x - y\|^2. \quad (1.21)$$

Agora, sabemos que

$$\alpha f(x) + (1 - \alpha)f(y) - \frac{\mu}{2}\alpha(1 - \alpha)\|x - y\|^2 - \frac{\mu}{2}\|\alpha x + (1 - \alpha)y\|^2 \quad (1.22)$$

$$= \alpha(f(x) - \frac{\mu}{2}\|x\|^2) + (1 - \alpha)(f(y) - \frac{\mu}{2}\|y\|^2) - \frac{\mu}{2}R, \quad (1.23)$$

onde $R := \alpha(1 - \alpha)\|x - y\|^2 + \|\alpha x + (1 - \alpha)y\|^2 - \alpha\|x\|^2 - (1 - \alpha)\|y\|^2$.

Usando que $\|\cdot\|^2 = \langle \cdot, \cdot \rangle$ verificamos diretamente que $R = 0$.

O resultado agora segue diretamente de (1.21), (1.22) e (1.23) e das definições de convexidade forte e convexidade. Se f é μ -fortemente convexa, usamos (1.21) em (1.22) e o fato de que (1.22) e (1.23) são iguais e com a definição de convexidade, vemos que $f - \frac{\mu}{2}\|\cdot\|^2$ é convexa. Por outro lado, supondo que $f - \frac{\mu}{2}\|\cdot\|^2$ é convexa, vemos que (1.23) majora $f(\alpha x + (1 - \alpha)y) - \frac{\mu}{2}\|\alpha x + (1 - \alpha)y\|^2$ e usando que (1.23) é igual a (1.22), vemos que f é μ -fortemente convexa. ■

Na sequência, nos referimos a funções convexas como funções 0-fortemente convexas. Esta terminologia tem o benefício de nos permitir tratar os casos convexo e fortemente convexo simultaneamente, de modo que podemos usar a definição de convexidade forte para $\mu \geq 0$.

Considere um espaço de dimensão finita \mathcal{X} com um produto interno denotado por $\langle \cdot, \cdot \rangle_{\mathcal{X}}$ e uma norma denotada por $\|\cdot\|_{\mathcal{X}}$ que não necessariamente é a norma induzida pelo produto interno. Considere o seguinte problema de otimização composta

$$\inf_{x \in \mathcal{X}} \{f(x) + g(x)\}, \quad (1.24)$$

onde $f : X \subset \mathcal{X} \rightarrow \mathbb{R}$ e $g : \mathcal{X} \rightarrow [-\infty, +\infty]$ satisfazem as seguintes condições:

A.1) g é uma função própria fechada μ -fortemente convexa;

A.2) X é um conjunto convexo tal que $X \supset \text{dom } g$;

A.3) existem uma constante $L > 0$ e uma função $\nabla f : X \rightarrow X$ (que não é necessariamente o gradiente de f apesar da notação) tais que para todos $x, x' \in X$,

$$f(x') + \langle \nabla f(x'), x - x' \rangle_{\mathcal{X}} \leq f(x) \leq f(x') + \langle \nabla f(x'), x - x' \rangle_{\mathcal{X}} + \frac{L}{2} \|x - x'\|_{\mathcal{X}}^2.$$

Mesmo que a aplicação ∇f não seja necessariamente o gradiente de f , como ela desempenha um papel similar ao do gradiente, denotá-la-emos desta forma.

O algoritmo acelerado para resolver o problema (1.24) requer a especificação de um ponto inicial $x_0 \in \text{dom } g$ e uma função $h : \mathcal{X} \rightarrow (-\infty, \infty]$ satisfazendo

A.4) h é uma função própria convexa e fechada tal que $\text{dom } h \supset \text{dom } g$;

A.5) h é 1-fortemente convexa em $\text{dom } g$;

A.6) $x_0 = \text{argmin}\{h(x) : x \in \text{dom } g\}$.

Vemos que, se $\text{dom } g$ é fechado, então como h é 1-fortemente convexa, e portanto estritamente convexa, segue que h é estritamente convexa em um conjunto fechado, e assim o problema em A.6 sempre tem um único minimizador global, o qual é no caso x_0 . O caso especial do algoritmo abaixo quando $\mu = 0$, coincide com a variante acelerada de Algoritmo 3 de (TSENG, 2008).

Algoritmo 1: Uma variante do algoritmo acelerado de Nesterov

Passo 0 sejam μ a constante de convexidade forte de g e L a constante satisfazendo A.3, e

faça $A_0 := 0$, $\Lambda_0 \equiv 0$, $k = 1$ e $\tilde{x}_0 = x_0$ onde x_0 é como em A.6;

Passo 1 compute

$$A_k := A_{k-1} + \frac{(1 + \mu A_{k-1}) + \sqrt{(1 + \mu A_{k-1})^2 + 4L(1 + \mu A_{k-1})A_{k-1}}}{2L}, \quad (1.25)$$

$$\begin{aligned} \check{x}_k &:= \frac{A_{k-1}}{A_k} \tilde{x}_{k-1} + \frac{A_k - A_{k-1}}{A_k} x_{k-1}, \\ \Lambda_k(\cdot) &:= \frac{A_{k-1}}{A_k} \Lambda_{k-1}(\cdot) + \frac{A_k - A_{k-1}}{A_k} [f(\check{x}_k) + \langle \nabla f(\check{x}_k), \cdot - \check{x}_k \rangle_{\mathcal{X}}]; \end{aligned} \quad (1.26)$$

Passo 2 determine x_k e \tilde{x}_k como

$$x_k := \operatorname{argmin} \left\{ \Lambda_k(x) + g(x) + \frac{1}{A_k} h(x) \right\}, \quad (1.27)$$

$$\tilde{x}_k := \frac{A_{k-1}}{A_k} \tilde{x}_{k-1} + \frac{A_k - A_{k-1}}{A_k} x_k; \quad (1.28)$$

Passo 3 faça $k \leftarrow k + 1$ e retorne ao Passo 1.

fim

Antes de estabelecermos o resultado de convergência do Algoritmo 1, precisamos do seguinte resultado auxiliar, que será mostrado mais adiante no Lema 4.1 da Seção 4.1.

Lema 1.1 *Considere U um subconjunto convexo de um espaço com produto interno \mathcal{Z} e $\psi, h : U \mapsto \mathbb{R}$ duas funções convexas, com h diferenciável. Denote por d dada por*

$$d(u_1, u_2) := d_{u_2}(u_1) = h(u_1) - h(u_2) - \langle \nabla h(u_2), u_1 - u_2 \rangle, \quad (1.29)$$

onde ∇h denota o gradiente da função h . Suponha que x é a solução ótima de

$$\min_{u \in U} \{ \psi(u) + \eta d_{u_0}(u) \},$$

para algum $u_0 \in U$ e $\eta > 0$. Então,

$$\psi(u) + \eta d_{u_0}(u) \geq \min_{u \in U} \{ \psi(u) + \eta d_{u_0}(u) \} + \eta d_x(u), \quad \forall u \in U.$$

Observamos que a função d dada por (1.29) é a distância de Bregman gerada por h . Esta noção será propriamente estabelecida na Seção 4.1, para efeito de prova desta proposição basta supor que d é dada por (1.29). O resultado principal que estabelece a convergência do algoritmo acelerado acima é o seguinte.

Proposição 1.12 *As sequências $\{A_k\}$, $\{\tilde{x}_k\}$ e $\{\Lambda_k\}$ geradas pelo Algoritmo 1 satisfazem, para todo $k \geq 1$, as seguintes desigualdades:*

$$A_k \geq \frac{1}{L} \max \left\{ \frac{k^2}{4}, \left(1 + \sqrt{\frac{\mu}{4L}} \right)^{2(k-1)} \right\}, \quad (1.30)$$

$$\sum_{i=1}^k \frac{A_i - A_{i-1}}{\sqrt{A_i}(1 + \mu A_{i-1})} \leq \frac{2k}{\sqrt{L}}, \quad (1.31)$$

$$\Lambda_k \leq f, \quad A_k(f + g)(\tilde{x}_k) \leq A_k(\Lambda_k(x) + g(x)) + [h(x) - h(x_0)], \quad \forall x \in \text{dom } g. \quad (1.32)$$

Primeiro damos duas provas alternativas de (1.32). Estas provas são adaptações das demonstrações encontradas em (TSENG, 2008) e (HE; MONTEIRO, 2014). Porém como estas usam notação e contexto muito diferentes, optamos por apresentar a adaptação aqui.

Prova. Primeira prova:

Denote $p := f + g$ e para todo $x, x' \in X$

$$l_f(x, x') := f(x') + \langle \nabla f(x'), x - x' \rangle.$$

Assim, por (1.26), temos que

$$\Lambda_k(x) = \sum_{i=1}^k \frac{A_i - A_{i-1}}{A_k} l_f(x, \tilde{x}_i). \quad (1.33)$$

Defina a sequência

$$\Theta_k := \min_{x \in X} A_k(\Lambda_k + g)(x) + h(x) - h(x_0).$$

Afirmamos que $\Theta_{k+1} - \Theta_k \geq A_{k+1}p(\tilde{x}_{k+1}) - A_k p(\tilde{x}_k)$. Com efeito, como h é 1-fortemente convexa e g é μ -fortemente convexa, temos que a função a ser minimizada na definição de

Θ_k é $1 + \mu A_k$ -fortemente convexa. Assim, pelo Lema 1.1 e (1.27), temos que para todo $x \in X$,

$$\Theta_k + \frac{1 + \mu A_k}{2} \|x_k - x\|^2 \leq A_k(\Lambda_k + g)(x) + h(x) - h(x_0).$$

Em particular, fazendo $x = x_{k+1}$ e usando (1.27) e (1.33), temos que

$$\begin{aligned} \Theta_k + \frac{1 + \mu A_k}{2} \|x_{k+1} - x_k\|^2 &\leq A_k(\Lambda_k + g)(x_{k+1}) + h(x_{k+1}) - h(x_0) = \\ &= \Theta_{k+1} - (A_{k+1} - A_k)l_f(x_{k+1}, \check{x}_{k+1}). \end{aligned} \quad (1.34)$$

Por outro lado, pela convexidade de $l_f(\cdot, \check{x}_{k+1})$, (1.28) e A.3, temos que

$$\begin{aligned} A_{k+1} [l_f(\tilde{x}_{k+1}, \check{x}_{k+1}) + g(\tilde{x}_{k+1})] &\leq (A_{k+1} - A_k) [l_f(x_{k+1}, \check{x}_{k+1}) + g(x_{k+1})] + \\ A_k [l_f(\tilde{x}_k, \check{x}_{k+1}) + g(\tilde{x}_k)] &\leq (A_{k+1} - A_k) [l_f(x_{k+1}, \check{x}_{k+1}) + g(x_{k+1})] + A_k p(\tilde{x}_k). \end{aligned} \quad (1.35)$$

Agora note pela definição de A_{k+1} que a mesma é a solução da equação quadrática (em A_{k+1})

$$L(A_{k+1} - A_k)^2 = A_{k+1}(1 + \mu A_{k+1}), \quad (1.36)$$

implicando (com as definições de \tilde{x}_k e x_k dadas em (1.28) e (1.27), respectivamente)

$$\|\tilde{x}_{k+1} - \check{x}_{k+1}\|^2 = \left\| \frac{A_{k+1} - A_k}{A_{k+1}} (x_{k+1} - x_k) \right\|^2 = \frac{1 + \mu A_k}{L A_{k+1}} \|x_{k+1} - x_k\|^2. \quad (1.37)$$

Usando (1.35) e (1.37) na desigualdade (1.34), temos que

$$\begin{aligned} \Theta_{k+1} - \Theta_k &\geq \frac{1 + \mu A_k}{2} \|x_{k+1} - x_k\|^2 + A_{k+1} [l_f(\tilde{x}_{k+1}, \check{x}_{k+1}) + g(\tilde{x}_{k+1})] - A_k p(\tilde{x}_k) \\ &\geq \frac{L A_{k+1}}{2} \|\tilde{x}_{k+1} - \check{x}_{k+1}\|^2 - A_k p(\tilde{x}_k) + A_{k+1} [l_f(\tilde{x}_{k+1}, \check{x}_{k+1}) + g(\tilde{x}_{k+1})] \geq \\ &A_{k+1} p(\tilde{x}_{k+1}) - A_k p(\tilde{x}_k), \end{aligned}$$

onde na última desigualdade usamos A.3.

Sendo assim provamos a afirmação. Segue que a sequência $\Theta_k - A_k p(\hat{x}_{k+1})$ é não decrescente, e usando que $A_0 = 0$ e a definição de Θ_k , vemos que

$$\Theta_k - A_k p(\hat{x}_k) \geq \Theta_0 = \min_{x \in X} [h(x) - h(x_0)] \geq 0.$$

Finalmente, usando a definição de Θ_k e a última desigualdade, conseguimos a segunda desigualdade em (1.32). A primeira desigualdade segue de (1.26), da convexidade de f , e do fato de que uma combinação convexa de funções minorizando f resulta em uma função minorizando f .

Segunda prova:

Denote p , $l_f(x, x')$ e a sequência Θ_k conforme a primeira prova. Esta prova será por indução. Para $k = 0$ a desigualdade (1.32) é trivial. Suponha que

$$A_k p(\tilde{x}_k) \leq \Theta_k. \quad (1.38)$$

Vamos mostrar que $A_{k+1} p(\tilde{x}_{k+1}) \leq \Theta_{k+1}$. Usando (1.38) e A.3 temos que

$$\begin{aligned} & A_k(\Lambda_k + g)(x) + (A_{k+1} - A_k) [l_f(x, \check{x}_{k+1}) + g(x)] + h(x) - h(x_0) \\ & \geq A_k p(\tilde{x}_k) + \frac{1 + \mu A_k}{2} \|x - x_k\|^2 + (A_{k+1} - A_k) [l_f(x, \check{x}_{k+1}) + g(x)] \\ & \geq A_k [l_f(\tilde{x}_k, \check{x}_{k+1}) + g(\tilde{x}_k)] + (A_{k+1} - A_k) [l_f(x, \check{x}_{k+1}) + g(x)] + \frac{1 + \mu A_k}{2} \|x - x_k\|^2, \end{aligned}$$

a qual é verdadeira para todo $x \in X$. Pela definição de Θ_{k+1} temos que

$$\Theta_{k+1} \geq A_k [l_f(\tilde{x}_k, \check{x}_{k+1}) + g(\tilde{x}_k)] + (A_{k+1} - A_k) [l_f(x, \check{x}_{k+1}) + g(x)] + \frac{1 + \mu A_k}{2} \|x - x_k\|^2. \quad (1.39)$$

Agora, note que, como observado na primeira prova, A_{k+1} satisfaz (1.36). Implicando (com as definições de \hat{x}_k e x_k em (1.28) e (1.27), respectivamente) (1.37).

Fazendo $x = x_{k+1}$ em (1.39) e usando a convexidade de $l_f(\cdot, \check{x}_{k+1})$, (1.28), (1.37) e A.3, temos que

$$\Theta_{k+1} \geq A_{k+1} [l_f(\tilde{x}_{k+1}, \check{x}_{k+1}) + g(\tilde{x}_{k+1})] + \frac{LA_{k+1}}{2} \|\tilde{x}_{k+1} - \check{x}_{k+1}\|^2 \geq A_{k+1} p(\tilde{x}_{k+1}), \quad (1.40)$$

o que prova (1.32).

Agora vamos provar (1.30): Fixe $k \geq 1$. Usando a relação (1.36) e o fato de que $A_k \leq A_{k+1}$, vemos que

$$A_k(1 + \mu A_{k-1}) = L(A_k - A_{k-1})^2 = L(A_k^{\frac{1}{2}} - A_{k-1}^{\frac{1}{2}})^2 (A_k^{\frac{1}{2}} + A_{k-1}^{\frac{1}{2}})^2 \leq 4LA_k(A_k^{\frac{1}{2}} - A_{k-1}^{\frac{1}{2}})^2,$$

implicando

$$\begin{aligned} A_k &\geq \left(A_{k-1}^{\frac{1}{2}} + \sqrt{\frac{\mu A_{k-1} + 1}{4L}} \right)^2 = \\ &A_{k-1} + A_{k-1}^{\frac{1}{2}} \sqrt{\frac{1 + \mu A_{k-1}}{L}} + \frac{1 + \mu A_{k-1}}{4L}. \end{aligned}$$

Usando um argumento usual de indução, vemos que a última relação implica

$$A_k \geq \frac{k^2(1 + \mu A_{k-1})}{4L}. \quad (1.41)$$

Por outro lado, a definição de A_k dada no Algoritmo 1 implica

$$\begin{aligned} A_k &= A_{k-1} + \frac{1 + \mu A_{k-1}}{2L} + \frac{\sqrt{(1 + \mu A_{k-1})^2 + 4L(1 + \mu A_{k-1})A_{k-1}}}{2L} \\ &\geq A_{k-1} + \frac{1 + \mu A_{k-1}}{4L} + \frac{\sqrt{LA_{k-1}(1 + \mu A_{k-1})}}{L} = \left(\sqrt{A_{k-1}} + \frac{1}{2L} \sqrt{L(1 + \mu A_{k-1})} \right)^2 \\ &\geq \left(\sqrt{A_{k-1}} + \frac{1}{2L} \sqrt{L\mu A_{k-1}} \right)^2 = A_{k-1} \left(1 + \frac{1}{2L} \sqrt{\mu L} \right)^2, \end{aligned}$$

a qual também pode ser verificada a implicar por indução

$$A_k \geq \left(1 + \sqrt{\frac{\mu}{4L}} \right)^{2(k-1)}. \quad (1.42)$$

Finalmente, usando (1.41) e (1.42) e o fato de que $A_k \geq A_1 = 1/L$, conseguimos (1.30).

E agora provaremos (1.31). Fixado $k \geq 0$, para cada $i \geq k$, temos, por (1.25) que

$$\begin{aligned} A_i &= A_{i-1} + \frac{1 + \mu A_{i-1}}{2L} + \frac{\sqrt{1 + 4L(1 + \mu A_{i-1})A_{i-1}}}{2L} \\ &\leq A_{i-1} + \frac{1 + \mu A_{i-1}}{L} + \frac{2\sqrt{A_{i-1}(1 + \mu A_{i-1})}}{\sqrt{2L}} \leq \left(\sqrt{A_{i-1}} + \frac{1 + \mu A_{i-1}}{\sqrt{L}} \right)^2. \end{aligned}$$

Portanto segue que

$$\sqrt{A_i} - \sqrt{A_{i-1}} \leq \frac{1 + \mu A_{i-1}}{\sqrt{L}},$$

o que implica

$$\sum_{i=1}^k \frac{A_i - A_{i-1}}{\sqrt{A_i}(1 + \mu A_{i-1})} = \sum_{i=1}^k \frac{(\sqrt{A_i} - \sqrt{A_{i-1}})(\sqrt{A_i} + \sqrt{A_{i-1}})}{\sqrt{A_i}(1 + \mu A_{i-1})} \leq \sum_{i=1}^k \frac{2(\sqrt{A_i} - \sqrt{A_{i-1}})}{1 + \mu A_{i-1}} \leq \frac{2k}{\sqrt{L}},$$

provando (1.31). ■

1.5 MÉTODO HPE PARA O PROBLEMA DE INCLUSÃO MONÓTONA

Nesta seção especificamos o método (Euclidiano) híbrido extragradiente proximal (HPE) para resolver um problema de inclusão monótona geral da forma (1.3) introduzido em (SOLODOV; SVAITER, 1999a). Conforme o nome sugere, usamos a distância Euclidiana para sua especificação. Apresentamos sua versão não-Euclidiana no Capítulo 4. Durante esta seção consideramos o problema de encontrar um zero de $T : \mathcal{Z} \rightrightarrows \mathcal{Z}$, onde \mathcal{Z} é um espaço com produto interno.

A seguir formulamos o método (Euclidiano) híbrido extragradiente proximal.

Algoritmo 2: Método (Euclidiano) híbrido extragradiente proximal (HPE)

Passo 0 tome $z_0 \in \mathcal{Z}$ e faça $j = 1$;

Passo 1 escolha $\sigma_j \in [0, 1]$, encontre $\lambda_j > 0$ e $(\tilde{z}_j, z_j, \varepsilon_j) \in \mathcal{Z} \times \mathcal{Z} \times \mathbb{R}_+$ tais que

$$r_j := \frac{1}{\lambda_j}(z_{j-1} - z_j) \in T^{\varepsilon_j}(\tilde{z}_j), \quad (1.43)$$

$$\|\tilde{z}_j - z_j\|^2 + 2\lambda_j\varepsilon_j \leq \sigma_j\|\tilde{z}_j - z_{j-1}\|^2; \quad (1.44)$$

Passo 2 faça $j \leftarrow j + 1$ e retorne ao Passo 1.

fm

Note que o Algoritmo 2 não especifica como determinar λ_j e $(\tilde{z}_j, z_j, \varepsilon_j)$ satisfazendo (1.43) e (1.44). O modo de determinar λ_j e $(\tilde{z}_j, z_j, \varepsilon_j)$ dependerá da versão do método em consideração e das propriedades do operador T . Note também que se $\sigma_j = 0$, então (1.44) implica $\varepsilon_j = 0$ e $z_j = \tilde{z}_j$, e portanto $r_j \in T(z_j)$, por causa de (1.43).

O próximo lema estabelece limites de complexidade para o Algoritmo 2. Sua prova pode ser encontrada no Teorema 4.4 de (MONTEIRO; SVAITER, 2010b).

Lema 1.2 *Considere $\{\tilde{z}_j\}$, $\{r_j\}$ e $\{\varepsilon_j\}$ sequências geradas pelo Algoritmo 2. Considere d_0 a distância (Euclidiana) de z_0 ao conjunto de soluções do problema de inclusão (1.3), então para todo $j \in \mathbb{N}$ vale $r_j \in T^{\varepsilon_j}(\tilde{x}_j)$ e existe $i \leq j$ tal que*

$$\|r_i\| \leq d_0 \sqrt{\frac{1+\sigma}{1-\sigma} \left(\frac{1}{\sum_{s=1}^j \lambda_s^2} \right)}, \quad \varepsilon_i \leq \frac{\sigma^2 d_0^2}{2(1-\sigma^2) \sum_{s=1}^j \lambda_s^2}. \quad (1.45)$$

1.6 TRABALHOS RELACIONADOS

Nesta seção apresentamos uma breve descrição dos trabalhos recentes relacionados com os problemas em consideração.

No contexto de desigualdades variacionais, Nemirovski (NEMIROVSKI, 2004) estabeleceu complexidade ergódica para uma extensão do método de Korpelevich (KORPELEVIČ, 1976), mais precisamente, o algoritmo mirror-prox, supondo que o conjunto viável do problema é limitado. Mais recentemente, Dang e Lan (DANG; LAN, 2015) estabeleceram resultados de complexidade para uma classe de métodos não-Euclidianos extragradientes para resolver desigualdades variacionais, quando o operador não necessariamente é monótono. Ainda, Chen et al. (CHEN; LAN; OUYANG, 2014) estabeleceram resultados de complexidade para um método mirror-prox acelerado que determina soluções fracas de uma classe de problemas de desigualdade variacional. Eles obtiveram complexidade ótima para o caso em que o conjunto viável do problema é limitado.

Nesterov ([NESTEROV, 2005](#)) desenvolveu um algoritmo de suavização para resolver problemas de ponto-de-sela de estrutura bilinear, supondo que X e Y são conjuntos convexos e compactos. O esquema consiste em aproximar a função objetivo do problema de ponto-de-sela associado por uma função convexa diferenciável de gradiente Lipschitziano e aplicar um método de gradiente acelerado (e.g. ([NESTEROV, 2005](#); [TSENG, 2008](#))) para o problema aproximado resultante.

O método HPE e seus resultados de convergência são estudados em ([SOLODOV; SVAITER, 1999a](#)) e posteriormente em ([SOLODOV; SVAITER, 2000](#)), e sua complexidade foi estabelecida em ([MONTEIRO; SVAITER, 2010b](#)) (alternativamente ([MONTEIRO; SVAITER, 2011](#); [MONTEIRO; SVAITER, 2013](#))). Os resultados de complexidade obtidos em ([MONTEIRO; SVAITER, 2010b](#)) dependem da distância do ponto inicial à solução ao invés do diâmetro do conjunto viável. Aplicações do método HPE para vários algoritmos de ordem zero (resp., primeira ordem) para resolver problemas de desigualdade variacional e inclusões monótonas (resp., problemas de ponto-de-sela) são discutidos em ([MONTEIRO; SVAITER, 2010b](#)) e ([MONTEIRO; SVAITER, 2011](#); [MONTEIRO; SVAITER, 2013](#)). Mais especificamente, interpretando o método de Korpelevich ([KORPELEVIČ, 1976](#)) e o método modificado de Tseng (MF-BS) ([TSENG, 2000](#)) como casos especiais do método HPE, os autores estabeleceram em ([MONTEIRO; SVAITER, 2011](#); [MONTEIRO; SVAITER, 2010b](#)) complexidade ergódica e pontual para estes métodos aplicados a: desigualdades variacionais monótonas, problemas de inclusão monótona consistindo da soma de um operador Lipschitziano e um operador monótono maximal de resolvente facilmente computável e problemas de ponto-de-sela.

Solodov e Svaiter ([SOLODOV; SVAITER, 2000](#)) propuseram outras versões do método NE-HPE, que permite determinar soluções aproximadas de (1.3) de acordo com

(1.6) no caso em que $\varepsilon = 0$. Finalmente, extensões do método proximal para o contexto de distâncias de Bregman foram estudados em (CENSOR; ZENIOS, 1992; CHEN; TEBoulle, 1993; ECKSTEIN, 1993; EGGERMONT, 1990; IUSEM; SOLODOV, 1997; KIWIEL, 1997), dentre outros. No entanto, nenhum destes trabalhos apresenta resultados de complexidade.

2 UMA VERSÃO ACELERADA DO ALGORITMO HPE PARA O PROBLEMA DE PONTO-DE-SELA

Neste capítulo apresentamos uma de nossas contribuições, que é o algoritmo HPE para resolver o problema de ponto-de-sela (SP-HPE), o qual consiste em aplicar o método HPE para o problema de inclusão monótona (1.3). Veremos que o algoritmo SP-HPE faz uso de um subproblema que depende da versão específica do método para ser resolvido. Sendo assim, na Seção 2.2 tratamos deste assunto descrevendo um caso particular do algoritmo SP-HPE, quando se tem que a função de duas variáveis em questão possui uma estrutura específica.

Durante este capítulo, supomos que \mathcal{X} , \mathcal{Y} , \mathcal{Z} , X , Y , Z , $\langle \cdot, \cdot \rangle_{\mathcal{X}}$, $\langle \cdot, \cdot \rangle_{\mathcal{Y}}$, $\langle \cdot, \cdot \rangle$ e $\hat{\Phi}$ são como na Subseção 1.3.1. Além disso, denote por $\| \cdot \|_{\mathcal{X}}$ e $\| \cdot \|_{\mathcal{Y}}$ respectivamente as normas em \mathcal{X} e \mathcal{Y} induzidas pelos correspondentes produtos internos. Nosso problema de interesse neste capítulo é o problema de ponto-de-sela $\text{SP}(\hat{\Phi}; Z)$, definido na Subseção 1.3.1, munido de uma certa estrutura no espaço \mathcal{X} , que consiste na existência de uma função própria convexa e fechada $\phi : \mathcal{X} \rightarrow (-\infty, +\infty]$ e uma função $\Phi : \text{dom } \Phi \supset Z \rightarrow \mathbb{R}$ satisfazendo

$$\text{dom } \phi = X, \tag{2.1}$$

$$\hat{\Phi}(x, y) = \Phi(x, y) + \phi(x), \quad \forall (x, y) \in Z, \tag{2.2}$$

e as seguintes condições adicionais:

B.1) Z é um conjunto convexo limitado e não-vazio;

B.2) para todo $z \in Z$, a função $\hat{\Phi}_z$ dada por (1.14) é fechada e convexa;

B.3) para todo $y \in Y$, a função $\Phi(\cdot, y)$ é diferenciável em X e existem constantes não

negativas L_{xx} e L_{xy} tais que

$$\|\nabla_x \Phi(x', y') - \nabla_x \Phi(x, y)\|_{\mathcal{X}}^* \leq L_{xx}\|x - x'\|_{\mathcal{X}} + L_{xy}\|y - y'\|_{\mathcal{Y}}, \quad \forall (x, y), (x', y') \in X \times Y,$$

onde $\|\cdot\|_{\mathcal{X}}^*$ denota a norma dual $\|\cdot\|_{\mathcal{X}}$ definida como

$$\|x\|_{\mathcal{X}}^* := \max_{\|x'\|_{\mathcal{X}}=1} \{\langle x, x' \rangle_{\mathcal{X}} : x' \in \mathcal{X}\}, \quad \forall x \in \mathcal{X}.$$

Observe que B.2 e a Proposição 1.9 (b) implicam o fato de que o operador $T_{\hat{\Phi}}$ dado por (1.15) é monótono maximal.

Nosso objetivo neste capítulo é desenvolver uma versão acelerada do método HPE para resolver (aproximadamente) o problema de ponto-de-sela $SP(\hat{\Phi}; Z)$, ou equivalentemente, a inclusão (1.16), com as hipóteses mencionadas acima. A solução aproximada considerada será no sentido da Definição 1.1. Na próxima seção descrevemos o algoritmo SP-HPE em sua forma geral e derivamos posteriormente, na Seção 2.2, uma versão acelerada.

2.1 O ALGORITMO SP-HPE PARA O PROBLEMA DE PONTO-DE-SELA

Nesta seção apresentamos o algoritmo SP-HPE, que consiste em especializar as condições de HPE (1.5) para o contexto do problema de ponto-de-sela aplicando-as ao operador monótono definido em (1.15).

Começamos enunciando o método HPE especializado para resolver o problema de ponto-de-sela de $SP(\hat{\Phi}; Z)$.

Algoritmo 3: Método híbrido extragradiente proximal para $SP(\hat{\Phi}; Z)$

Passo 0 dados $z_0 := (x_0, y_0) \in \mathcal{Z}$, $\lambda > 0$ e $0 \leq \sigma < 1$, faça $k = 1$;

Passo 1 encontre $\tilde{z}_k := (\tilde{x}_k, \tilde{y}_k) \in \mathcal{Z}$, $r_k = (r_k^x, r_k^y) \in \mathcal{Z}$ e $\varepsilon_k \geq 0$ tais que

$$r_k \in \partial_{\varepsilon_k} [\hat{\Phi}(\cdot, \tilde{y}_k) - \hat{\Phi}(\tilde{x}_k, \cdot)](\tilde{z}_k), \quad \text{com } \hat{\Phi} \text{ dada por (2.2),} \quad (2.3)$$

$$\|\lambda r_k^x + \tilde{x}_k - x_{k-1}\|_{\mathcal{X}}^2 + \|\lambda r_k^y + \tilde{y}_k - y_{k-1}\|_{\mathcal{Y}}^2 + 2\lambda\varepsilon_k \leq \sigma^2 \left(\|\tilde{x}_k - x_{k-1}\|_{\mathcal{X}}^2 + \|\tilde{y}_k - y_{k-1}\|_{\mathcal{Y}}^2 \right); \quad (2.4)$$

Passo 2 faça $x_k = x_{k-1} - \lambda r_k^x$, $y_k = y_{k-1} - \lambda r_k^y$ e $k \leftarrow k + 1$, e retorne ao Passo 1.

fim

Agora fazemos uma série de observações acerca do Algoritmo 3. Primeiramente, uma vez que o ε -subdiferencial está sempre contido no ε -enlargamento (ver Proposição 1.6), o Algoritmo 3 é um caso especial do método HPE no qual $\lambda_k := \lambda$. Na realidade, o Algoritmo 3 poderia ser enunciado em termos de uma seqüência de tamanhos de passo variável $\{\lambda_k\}$; contudo para efeito de simplicidade fazemos $\lambda_k = \lambda$. Em segundo lugar, similarmente ao método HPE, o Algoritmo 3 não especifica como determinar \tilde{z}_k , r_k e ε_k satisfazendo as condições de HPE em (2.3) e (2.4). Isto será explorado adiante, na Subseção 2.2.1.

O resultado seguinte, já conhecido, descreve as propriedades de taxa de convergência ergódica e pontual do Algoritmo 3.

Teorema 2.1 *Considere as seqüências $\{\tilde{z}_k\}, \{r_k\}$ e $\{\varepsilon_k\}$ geradas pelo Algoritmo 3 e, para todo $k \geq 0$, defina*

$$\tilde{z}_k^a = (\tilde{x}_k^a, \tilde{y}_k^a) := \frac{1}{k} \sum_{i=1}^k (\tilde{x}_i, \tilde{y}_i), \quad r_k^a := \frac{1}{k} \sum_{i=1}^k (r_i^x, r_i^y)$$

e

$$\varepsilon_k^a := \frac{1}{k} \sum_{i=1}^k [\varepsilon_i + \langle (\tilde{x}_i - \tilde{x}_k^a, \tilde{y}_i - \tilde{y}_k^a), (r_i^x, r_i^y) \rangle].$$

Considere d_0 a distância (Euclidiana) de (x_0, y_0) ao conjunto de soluções de $SP(\hat{\Phi}; Z)$.

Então, para todo $k \geq 0$, valem as seguintes afirmações:

- (a) (taxa de convergência pontual) a tripla $(\tilde{z}_k, r_k, \varepsilon_k)$ é um $(\|r_k\|, \varepsilon_k)$ -ponto-de-sela (ver a Definição 1.1) de $\hat{\Phi}$, ou equivalentemente, vale (2.3) e existe um índice $i \leq k$ tal que

$$\|r_i\| \leq \frac{d_0}{\lambda} \sqrt{\frac{1+\sigma}{k(1-\sigma)}}, \quad \varepsilon_i \leq \frac{\sigma^2 d_0^2}{2k\lambda(1-\sigma^2)};$$

- (b) (taxa de convergência ergódica) a tripla $(\tilde{z}_k^a, r_k^a, \varepsilon_k^a)$ é um $(\|r_k^a\|, \varepsilon_k^a)$ -ponto-de-sela de $\hat{\Phi}$, ou equivalentemente

$$r_k^a \in \partial_{\varepsilon_k^a}(\hat{\Phi}(\cdot, \tilde{y}_k^a) - \hat{\Phi}(\tilde{x}_k^a, \cdot))(\tilde{x}_k^a, \tilde{y}_k^a),$$

e

$$\|r_k^a\| \leq \frac{2d_0}{\lambda k}, \quad 0 \leq \varepsilon_k^a \leq \frac{2d_0^2}{\lambda k} \left(1 + \frac{\sigma}{\sqrt{1-\sigma^2}} \right).$$

Prova. A prova deste teorema se baseia na teoria desenvolvida em (MONTEIRO; SVAITER, 2013) para o caso mais geral de ε -enlargamentos. Uma prova para o caso específico da função $\hat{\Phi}(\cdot, \tilde{y}_k^a) - \hat{\Phi}(\tilde{x}_k^a, \cdot)$ pode ser encontrada no Teorema 3.4 de (HE; MONTEIRO, 2014). ■

2.2 UMA VERSÃO ACELERADA DO MÉTODO HPE PARA O PROBLEMA DE PONTO-DE-SELA

Nesta seção apresentamos uma versão do Algoritmo 3 descrito na Seção 2.1 onde usamos o método acelerado apresentado na Seção 1.4 para resolver o subproblema, dado pelas relações (2.3) e (2.4), em cada iteração do Algoritmo 3. Na Subseção 2.2.1 nos concentramos no algoritmo para resolver o subproblema, e também enunciamos suas propriedades de convergência. Na Seção 2.3 tratamos da complexidade do método.

2.2.1 Resolvendo o Subproblema de HPE

Nesta subseção apresentamos um esquema para determinar uma tripla $(\tilde{z}, r, \varepsilon)$ satisfazendo as condições de HPE (2.3)-(2.4) com $\hat{\Phi}$ satisfazendo B.1-B.3. O esquema é baseado na variante acelerada de Nesterov descrita na Seção 1.4 aplicada a um problema de ponto-de-sela associado que determina um ε -ponto-de-sela a cada iteração, no sentido da Definição 1.1.

Considere o seguinte problema correspondendo ao caso do Passo 1 do Algoritmo 3 quando λ é escolhido igual a 1.

(P1) *Considere conjuntos convexos $X \subset \mathcal{X}$ e $Y \subset \mathcal{Y}$, uma função $\hat{\Phi}$ satisfazendo B.1-B.3, um par $(x_-, y_-) \in \mathcal{Z}$ e um escalar $\sigma > 0$. O problema é determinar $\tilde{z} = (\tilde{x}, \tilde{y}) \in \mathcal{Z}$, $r = (r^x, r^y) \in \mathcal{X} \times \mathcal{Y}$ e $\tilde{\varepsilon} \geq 0$ tais que*

$$r \in \partial_{\tilde{\varepsilon}} \left[\hat{\Phi}(\cdot, \tilde{y}) - \hat{\Phi}(\tilde{x}, \cdot) \right] (\tilde{z}), \quad (2.5)$$

$$\|r^x + \tilde{x}_k - x_{k-1}\|_{\mathcal{X}}^2 + \|r^y + \tilde{y}_k - y_{k-1}\|_{\mathcal{Y}}^2 + 2\tilde{\varepsilon} \leq \sigma^2 \left(\|\tilde{x} - x_-\|_{\mathcal{X}}^2 + \|\tilde{y} - y_-\|_{\mathcal{Y}}^2 \right). \quad (2.6)$$

Apresentamos um esquema baseado na variante acelerada de Nesterov da Seção 1.4 para resolver (P1), com $\hat{\Phi}$ satisfazendo B.1-B.3.

A seguir fazemos duas observações acerca do problema (P1).

Primeiro, determinar uma solução exata de (P1), i.e., com $\sigma = 0$, é equivalente a determinar um ponto-de-sela de

$$\min_{x \in X} \max_{y \in Y} \hat{\Phi}_0(x, y) := \hat{\Phi}(x, y) + \frac{1}{2} \|x - x_-\|^2 - \frac{1}{2} \|y - y_-\|^2, \quad (2.7)$$

onde $\hat{\Phi}$ é dada por (2.2). Mais especificamente, se (\tilde{x}, \tilde{y}) é um ponto-de-sela exato do problema (2.7), então (\tilde{x}, \tilde{y}) , $(\tilde{r}^x, \tilde{r}^y) := (x_- - \tilde{x}, y_- - \tilde{y})$ e $\tilde{\varepsilon} := 0$ satisfazem (2.5) e (2.6) com $\sigma = 0$. Segundo, embora o problema de ponto-de-sela (2.7) tenha essencialmente a

mesma estrutura que aquele que estamos interessados em resolver, como veremos mais adiante, sua função primal tem a propriedade de que é a soma de uma função convexa g_0 de resolvente facilmente computável, e uma função f_0 convexa diferenciável e com gradiente Lipschitz-contínuo. Portanto, podemos obter soluções aproximadas de (2.7) usando a variante acelerada de Nesterov para problemas de otimização convexa composta como desenvolvido na Seção 1.4.

Em vista das observações acima, é razoável esperar que soluções aproximadas de (2.7) também sejam soluções de (P1) (com $\sigma > 0$).

A seguir, no resto da seção, nos concentramos em apresentar um esquema baseado na variante acelerada de Nesterov da Seção 1.4, aplicado a (2.7), para obter uma solução de (P1) e apresentar seu correspondente resultado de convergência.

Claramente (2.7) é um caso especial de (1.24), com o par (f, g) dado por

$$g(x) := \phi(x) + \frac{1}{2}\|x - x_-\|_{\mathcal{X}}^2, \quad (2.8)$$

$$f(x) := \max_{y \in Y} \left\{ \Phi_0(x, y) := \Phi(x, y) - \frac{1}{2}\|y - y_-\|^2 \right\}. \quad (2.9)$$

No que segue discutimos como o problema de ponto-de-sela regularizado (2.7) pode ser visto como um problema de otimização convexa composta da forma (1.24) satisfazendo as condições A.1-A.3. Primeiro precisamos do seguinte resultado clássico sobre a norma Euclidiana:

Lema 2.1 *Considere $m \geq 0$, a função $f : \mathcal{Z} \mapsto \mathbb{R}$ dada por $f(x) = m\|x\|^2$ é m -fortemente convexa.*

Prova. Para $m = 0$, o resultado é trivial. Tomemos $\mu > 0$. A ideia é usar o fato de que a norma Euclidiana vem de um produto interno. Com isso em mente, dado $t \in (0, 1)$, temos

que

$$\begin{aligned}
& f(tx + (1-t)y) - t(f(x) - (1-t)f(y)) = \\
& mt^2||x||^2 + 2mt(1-t)\langle x, y \rangle + m(1-t)^2||y||^2 - mt||x||^2 - m(1-t)||y||^2 = \\
& mt(t-1)||x||^2 + m(1-t)(1-t-1)||y||^2 + 2mt(1-t)\langle x, y \rangle = \\
& -mt(1-t)||x||^2 + 2mt(1-t)\langle x, y \rangle - mt(1-t)||y||^2 = \\
& -mt(1-t)(||x||^2 - 2\langle x, y \rangle + ||y||^2) = -mt(1-t)||x - y||^2 \\
& \leq -\frac{1}{2}mt(1-t)||x - y||^2.
\end{aligned}$$

Uma vez que $mt(1-t)||x - y||^2 \geq 0$, segue o resultado. ■

Pelas hipóteses e pelo Lema 2.1, é claro que a função g em (2.8) acima satisfaz A.1-A.2 com $\mu = 1$. O seguinte resultado implica que a função f em (2.9) acima satisfaz a condição A.3. Este resultado é um caso particular da Proposição 4.1 de (MONTEIRO; SVAITER, 2010a). A prova é feita de maneira totalmente análoga, mas exibimos a prova por conveniência.

Proposição 2.1 *Considere a função f dada por (2.9). Então a constante L_λ e a função $\nabla f : X \rightarrow X$ definidas, respectivamente, como*

$$L_\lambda := 2 \left(L_{xx} + L_{xy}^2 \right), \quad \nabla f(x) := \nabla_x \Phi(x, y(x)), \quad \forall x \in X, \quad (2.10)$$

onde L_{xx} e L_{xy} são as constantes que satisfazem B.3 e $y(x)$ é definido como

$$y(x) := \operatorname{argmax}_{y \in Y} \left\{ \Phi(x, y) - \frac{1}{2}||y - y_-||^2 \right\}, \quad \forall x \in X, \quad (2.11)$$

satisfazem a condição A.3.

Prova. Defina $\psi(x, y) := \Phi(x, y) - \frac{1}{2}||y - y_-||^2$. Fixe um ponto arbitrário $\hat{x} \in X$ e considere $y(\hat{x}) \in Y$ conforme (2.11), ou seja, temos que $f(\hat{x}) = \psi(\hat{x}, y(\hat{x}))$.

Primeiro vamos mostrar que $\nabla f(\hat{x}) := \nabla_x \Phi(\hat{x}, y(\hat{x})) \in \partial f(\hat{x})$.

De fato, para todo $x \in X$,

$$f(x) - f(\hat{x}) = f(x) - \psi(\hat{x}, y(\hat{x})) \geq \psi(x, y(\hat{x})) - \psi(\hat{x}, y(\hat{x})) \geq \langle \nabla f(\hat{x}), x - \hat{x} \rangle.$$

A igualdade vale por hipótese, a primeira desigualdade pela definição de f e a segunda pela convexidade de $\Phi(\cdot, y)$ e pelo fato de que $\nabla_x \psi(\hat{x}, y) = \nabla_x \Phi(\hat{x}, y)$ para todo y . Segue que $\nabla f(\hat{x}) \in \partial f(\hat{x})$.

Agora, para todo (x, y) vale $\nabla_x \psi(x, y) = \nabla_x \Phi(x, y)$, usando a convexidade de $\psi(\cdot, y)$, a desigualdade de Cauchy-Schwarz e B.3, temos que para todo (x, y)

$$\begin{aligned} \psi(\hat{x}, y) - \psi(x, y) - \langle \nabla f(\hat{x}), \hat{x} - x \rangle &\geq \langle \nabla_x \psi(x, y) - \nabla_x \psi(\hat{x}, y(\hat{x})), \hat{x} - x \rangle \geq \\ &\geq -\|\nabla_x \psi(x, y) - \nabla_x \psi(\hat{x}, y(\hat{x}))\| \|\hat{x} - x\| \geq -(L_{xx}\|\hat{x} - x\| + L_{xy}\|y - y(\hat{x})\|) \|\hat{x} - x\|. \end{aligned} \quad (2.12)$$

Além disso, como $\psi(x, \cdot)$ é 1-fortemente côncava, temos que

$$-\psi(\hat{x}, y) \geq -f(\hat{x}) + \frac{1}{2}\|y - y(\hat{x})\|^2. \quad (2.13)$$

Somando as desigualdades (2.12) e (2.13), temos que

$$\begin{aligned} \psi(x, y) &\leq f(\hat{x}) + \langle \nabla f(\hat{x}), x - \hat{x} \rangle + L_{xx}\|x - \hat{x}\|^2 + L_{xy}\|y - y(\hat{x})\| \|\hat{x} - x\| - \frac{1}{2}\|y - y(\hat{x})\|^2 \\ &\leq f(\hat{x}) + \langle \nabla f(\hat{x}), x - \hat{x} \rangle + L_{xx}\|x - \hat{x}\|^2 + \max_{t \in \mathbb{R}} \left\{ tL_{xy}\|x - \hat{x}\| - \frac{1}{2}t^2 \right\} = \\ &= f(\hat{x}) + \langle \nabla f(\hat{x}), x - \hat{x} \rangle + \left(L_{xx} + \frac{L_{xy}^2}{2} \right) \|x - \hat{x}\|^2, \quad \forall y \in Y. \end{aligned}$$

Tomando o supremo sobre Y vemos que a desigualdade também vale para $f(x)$ no lado esquerdo.

Por esta última desigualdade e pelo fato de que $\nabla f(\hat{x}) \in \partial f(\hat{x})$, vemos que

$$f(\hat{x}) + \langle \nabla f(\hat{x}), x - \hat{x} \rangle \leq f(x) \leq f(\hat{x}) + \langle \nabla f(\hat{x}), x - \hat{x} \rangle + \left(L_{xx} + \frac{L_{xy}^2}{2} \right) \|x - \hat{x}\|^2,$$

completando a prova. ■

Agora estamos prontos para enunciar o algoritmo acelerado para resolver o problema

$$\min_{x \in \mathcal{X}} \{p(x) := f(x) + g(x)\}, \quad (2.14)$$

com f e g definidas em (2.8) e (2.9), respectivamente.

Algoritmo 4: Algoritmo acelerado para resolver (2.14):

Passo 0 dados Φ_0, f, g, L_λ como definidos, respectivamente, em (2.9), (2.8) e (2.10), uma

função $\Lambda_0 \equiv 0$, $(x_-, y_-) \in \mathcal{X} \times \mathcal{Y}$, $\sigma \in (0, 1)$, faça

$$A_0 = 0, \quad k = 1,$$

$$\tilde{x}_0 = x_0 := x_-, \check{x}_0 := P_X(x_-), \quad \tilde{y}_0 := 0;$$

Passo 1

$$\begin{aligned} A_k &= A_{k-1} + \frac{(A_{k-1} + 1) + \sqrt{(A_{k-1} + 1)^2 + 4L_\lambda(A_{k-1} + 1)A_{k-1}}}{2L_\lambda}, \\ \check{x}_k &= \frac{A_{k-1}}{A_k} \tilde{x}_{k-1} + \frac{A_k - A_{k-1}}{A_k} x_{k-1}, \end{aligned} \quad (2.15)$$

compute $y(\check{x}_k)$ de acordo com (2.11), e a função afim Λ_k como

$$\Lambda_k(\cdot) := \frac{A_{k-1}}{A_k} \Lambda_{k-1}(\cdot) + \frac{A_k - A_{k-1}}{A_k} [f(\check{x}_k) + \langle \nabla f(\check{x}_k), \cdot - \check{x}_k \rangle_{\mathcal{X}}];$$

Passo 2 compute

$$\begin{aligned} \tilde{y}_k &= \frac{A_{k-1}}{A_k} \tilde{y}_{k-1} + \frac{A_k - A_{k-1}}{A_k} y(\check{x}_k), \\ x_k &= \operatorname{argmin}_{x \in X} \left\{ \Lambda_k(x) + g(x) + \frac{1}{2A_k} \|x - x_-\|^2 \right\}, \\ \tilde{x}_k &= \frac{A_{k-1}}{A_k} \tilde{x}_{k-1} + \frac{A_k - A_{k-1}}{A_k} x_k; \end{aligned} \quad (2.16)$$

Passo 3 se $\frac{1}{A_k} \leq \sigma$, pare e retorne

$$r_k := \begin{pmatrix} r_k^x \\ r_k^y \end{pmatrix} = \begin{pmatrix} x_k - x_- \\ y(\tilde{x}_k) - y_- \end{pmatrix},$$

caso contrário faça $k := k + 1$ e retorne ao Passo 1.

fim

Observe que o Passo 3 não faz parte da variante de Nesterov descrita na Seção 1.4.

Na verdade, este passo é o critério de parada. No Capítulo 5 veremos que este critério implica que a condição (2.4) é satisfeita.

A seguinte proposição é o resultado principal de convergência. Uma vez que a prova desta proposição faz uso de diversos resultados auxiliares, optamos por primeiro terminar a discussão e deixar sua prova para a Seção 2.4.

Proposição 2.2 *Considere $\{\tilde{y}_k\}, \{\tilde{x}_k\}, \{r_x^k\}, \{r_y^k\}$ geradas pelo Algoritmo 4 e*

$$c_k = 1 + \frac{1}{A_k}.$$

Então para todo $k \geq 1$ existe um número $\varepsilon_k \leq \frac{1}{2A_k} \|\tilde{x}_k - x_0\|^2$ tal que

$$r^k := \begin{pmatrix} r_x^k \\ r_y^k \end{pmatrix} \in \partial_{\varepsilon_k} [\hat{\Phi}(\cdot, \tilde{y}_k) - \hat{\Phi}(\tilde{x}_k, \cdot)](\tilde{x}_k, \tilde{y}_k) + \begin{pmatrix} c_k(\tilde{x}_k - x_-) \\ \tilde{y}_k - y_- \end{pmatrix},$$

$$\|r^k\|^2 + 2\varepsilon_k \leq \hat{\sigma}_k \left(\|\tilde{x}_k - x_- \|^2 + \|\tilde{y}_k - y_- \|^2 \right),$$

com

$$\hat{\sigma}_k := \frac{1}{A_k}.$$

Baseado nesta Proposição e no fato de que $A_k \geq \frac{k^2}{4L}$, dado por (1.30), vemos que o Algoritmo 4 resolve (P1) em, no máximo $\mathcal{O}(\lceil \sqrt{L} \rceil)$ iterações.

De fato, pela definição de $\hat{\sigma}_k$, e pelo fato de que $A_k \geq \frac{k^2}{4L}$, temos que

$$\hat{\sigma}_k \leq \frac{4L}{k^2}.$$

Logo, o algoritmo termina em no máximo $k_0 = \lceil 4L\sigma^{-2} \rceil$ iterações. Com efeito, se assim não fosse, teríamos

$$\sigma < \hat{\sigma}_k \leq \frac{4L}{k_0^2} \leq \sigma,$$

que é uma contradição. (A primeira destas desigualdades se deve ao critério de parada no Passo 3 e à hipótese de que o algoritmo não terminou em k_0 iterações). Segue que o algoritmo termina em no máximo $\mathcal{O}(\lceil \sqrt{L} \rceil)$ iterações. E veremos na Seção 2.4 que a Proposição 2.2 implica que ele resolve (P1), conforme desejado.

2.3 COMPLEXIDADE DO ALGORITMO ACELERADO PARA $\text{SP}(\hat{\Phi}; Z)$

Nesta seção apresentamos resultados sobre a complexidade do método acelerado para o problema de ponto-de-sela.

O seguinte resultado mostra que é tão geral resolver o problema (P1) para $\lambda = 1$ quanto para um tamanho de passo qualquer $\lambda > 0$.

Proposição 2.3 *Considere $\lambda > 0$, Φ uma função satisfazendo B.1-B.3 e a k -ésima iteração do Algoritmo 3. Se $(x, y); (r^u, r^v) \in \mathcal{X} \times \mathcal{Y}$ e $\varepsilon \geq 0$ resolvem (P1) com dados de entrada $\hat{\Phi} = \lambda\Phi$, $(x_-, y_-) = (x_{k-1}, y_{k-1})$ e $\sigma > 0$, então*

$$(\tilde{x}_k, \tilde{y}_k) := (x, y), \quad (r_k^x, r_k^y) := \frac{1}{\lambda}(r^u, r^v), \quad \varepsilon_k := \frac{1}{\lambda}\varepsilon$$

satisfazem as condições de HPE (2.3) e (2.4) do Passo 1 do Algoritmo 3.

Prova. Segue imediatamente da identidade

$$\lambda \partial_\varepsilon [\hat{\Phi}(\cdot, y) - \hat{\Phi}(x, \cdot)](x, y) = \partial_{\lambda\varepsilon} [\lambda \hat{\Phi}(\cdot, y) - \lambda \hat{\Phi}(x, \cdot)](x, y),$$

a qual vale para todo $\varepsilon \geq 0$, $\lambda \geq 0$ e $(x, y) \in \mathcal{X} \times \mathcal{Y}$. ■

Em vista do resultado acima, vemos que podemos usar o Algoritmo 4 para resolver o subproblema inexato regularizado

$$\min_{x \in X} \max_{y \in Y} \lambda \hat{\Phi}(x, y) + \frac{1}{2} \|x - x_-\|^2 - \frac{1}{2} \|y - y_-\|^2.$$

Algoritmo 5: Método acelerado SP-HPE para o problema de ponto-de-sela

Passo 0 dados $(x_0, y_0) \in \mathcal{X} \times \mathcal{Y}$, $\lambda > 0$ e $0 < \sigma < 1$ faça $k = 1$;

Passo 1 invoque o Algoritmo 4 com entrada

$$\Lambda_k^0 = \lambda \Lambda_k^0, \quad g^0 = \lambda g^0, \quad (u_0, v_0) = (x_{k-1}, y_{k-1}), \quad L = \lambda L,$$

e faça

$$(\tilde{x}_k, \tilde{y}_k) := (x, y), \quad r_k = (r_k^x, r_k^y) := \frac{1}{\lambda} (r^u, r^v), \quad \varepsilon_k := \frac{1}{\lambda} \tilde{\varepsilon},$$

onde (x, y) , (r^u, r^v) e $\tilde{\varepsilon}$ são retornados pelo Algoritmo 4;

Passo 2 faça $x_k = x_{k-1} - \lambda r_k^x$, $y_k = y_{k-1} - \lambda r_k^y$, $k \leftarrow k + 1$, e retorne ao Passo 1.

fim

Em vista da Proposição 2.5 a seguir, o Algoritmo 5 satisfaz as condições (2.3) e (2.4) do Passo 1 do Algoritmo 3. Sendo assim, o Algoritmo 5 é claramente um caso especial do Algoritmo 3. Segue que as estimativas de taxa de convergência pontual e ergódica (externa) para o Algoritmo 5 acelerado são como descritas nas afirmações (a) e (b) do Teorema 2.1, respectivamente.

Teorema 2.2 *Suponha que valem A.1-A.2, e que o conjunto (convexo) de soluções de $SP(\Phi, Z)$ é não-vazio, e considere d_0 a distância (Euclidiana) do ponto inicial (x_0, y_0) do Algoritmo 5 a este conjunto. Considere as sequências $\{(\tilde{x}_k, \tilde{y}_k)\}$, $\{(r_k^x, r_k^y)\}$ e $\{\varepsilon_k\}$ geradas pelo Algoritmo 5 e as seqüências ergódicas respectivas $\{(\tilde{x}_k^a, \tilde{y}_k^a)\}$, $\{r_k^a\}$ e $\{\varepsilon_k^a\}$ definidas no Teorema 2.1. Então, valem as seguintes afirmações:*

(a) *para todo par de escalares positivos (ρ, ε) , existe*

$$k_0 = \mathcal{O} \left(\max \left\{ 1, \left\lceil \frac{d_0}{\lambda \rho} \right\rceil, \left\lceil \frac{d_0^2}{\lambda \varepsilon} \right\rceil \right\} \right)$$

tal que para todo $k \geq k_0$, a tripla $((\tilde{x}_k^a, \tilde{y}_k^a), r_k^a, \varepsilon_k^a)$ é um (ρ, ε) -ponto-de-sela de $SP(\Phi, Z)$;

(b) *cada iteração do Algoritmo 5 performa no máximo*

$$\mathcal{O} \left(\lceil \sqrt{\lambda L} \rceil \right)$$

iterações internas.

Como consequência, o Algoritmo 5 determina um (ρ, ε) -ponto-de-sela de $SP(\Phi, Z)$ performando no máximo

$$\mathcal{O} \left(\lceil \sqrt{\lambda L} \rceil \max \left\{ 1, \left\lceil \frac{d_0}{\lambda \rho} \right\rceil, \left\lceil \frac{d_0^2}{\lambda \varepsilon} \right\rceil \right\} \right)$$

iterações internas.

Prova. Uma vez que o Algoritmo 5 é uma versão especial do Algoritmo 3, (a) segue imediatamente do Teorema 2.1 (b). A afirmação (b) segue da Proposição 2.2 com $L = \lambda L$. A última afirmação segue diretamente de (a) e (b). ■

2.4 CONVERGÊNCIA DO ALGORITMO ACELERADO PARA $SP(\Phi, Z)$

Nesta seção provamos o resultado de convergência enunciado na Subseção 2.2.1.

Primeiro precisamos do seguinte resultado técnico que exhibe um limitante para a função Λ_0^k definida no Algoritmo 4.

Lema 2.2 *Para todo $(x, y) \in X \times Y$ faça*

$$\Phi_0(x, y) := \Phi(x, y) - \frac{1}{2}\|y - y_0\|^2$$

Então a função Λ_k definida no Algoritmo 4 satisfaz

$$\Lambda_k(x) \leq \Phi_0(x, \tilde{y}_k).$$

Prova. Observe que para todo $i \geq 0$ temos que

$$f(\check{x}_i) + \langle \nabla f_0(\check{x}_i), x - \check{x}_i \rangle = \Phi_0(\check{x}_i, y(\check{x}_i)) + \langle \nabla_x \Phi(\check{x}_i, y(\check{x}_i)), x - \check{x}_i \rangle \leq \Phi_0(x, y(\check{x}_i)),$$

onde a última desigualdade deve-se à convexidade de $\Phi_0(\cdot, y)$ e ao fato de que $\nabla_x \Phi_0(x, y) = \nabla_x \Phi(x, y)$, $\forall (x, y) \in X \times Y$.

Somando para $i = 1$ até k e usando a concavidade de $\Phi(x, \cdot)$, temos que

$$\begin{aligned} \sum_{i=1}^k \frac{A_i - A_{i-1}}{A_k} [f_0(\check{x}_i) + \langle \nabla f_0(\check{x}_i), x - \check{x}_i \rangle] &\leq \sum_{i=1}^k \frac{A_i - A_{i-1}}{A_k} \Phi_0(x, y(\check{x}_i)) \\ &\leq \Phi_0\left(x, \sum_{i=1}^k \frac{A_i - A_{i-1}}{A_k} y(\check{x}_i)\right), \end{aligned}$$

o que com a definição de \tilde{y}_k (em (2.16)) completa a prova. ■

Dada uma função própria, convexa e fechada h e uma quadrática q , o próximo resultado estabelece a relação do ε -subdiferencial da soma das funções $(h + q)(u)$ e a soma do gradiente de q e do ε -subdiferencial de h .

Proposição 2.4 *Suponha que h é uma função própria convexa e fechada e q é uma quadrática com Hessiana Q positiva definida, ambas definidas num espaço normado X , e*

suponha que (z, r, ε) satisfazem $r \in \partial_\varepsilon(h + q)(z)$. Defina

$$\begin{aligned}\bar{z} &= \operatorname{argmin}_u \{h(u) + q(u)\}, \quad \delta = Q(z - \bar{z}), \\ \bar{\varepsilon} &= \varepsilon - \langle r, \bar{z} - z \rangle - \frac{1}{2} \langle z - \bar{z}, Q(z - \bar{z}) \rangle.\end{aligned}$$

Então, temos que

$$\begin{aligned}\delta &\in \nabla q(z) + \partial_\varepsilon h(z), \\ \bar{\varepsilon} + \|Q^{-\frac{1}{2}}(r - \delta)\|^2 &= \varepsilon + \frac{1}{2} \langle r, Q^{-1}r \rangle.\end{aligned}\tag{2.17}$$

Prova. Pela hipótese de que $r \in \partial_\varepsilon(h + q)(z)$ temos que para todo u ,

$$h(u) - h(z) + q(u) - q(z) \geq \langle r, u - z \rangle - \varepsilon.\tag{2.18}$$

Uma vez que a quadrática coincide com sua expansão de Taylor de segunda ordem, podemos dizer que

$$q(u) - q(z) = \langle \nabla q(z), u - z \rangle + \frac{1}{2} \langle u - z, Q(u - z) \rangle.\tag{2.19}$$

Usando (2.18) e (2.19), temos que para todo u ,

$$\frac{1}{2} \langle u - z, Q(u - z) \rangle + \langle \nabla q(z), u - z \rangle + h(u) - h(z) \geq -\varepsilon + \langle r, u - z \rangle.$$

Em particular, notando que o mínimo do lado esquerdo é \bar{z} , temos que

$$\min_u \frac{1}{2} \langle u - z, Q(u - z) \rangle + \langle \nabla q(z), u - z \rangle + h(u) - h(z) \geq -\varepsilon + \langle r, \bar{z} - z \rangle.$$

Uma vez que o lado direito está fixado, podemos usar o Lema 1.1 e obter

$$h(u) - h(z) \geq -\langle \nabla q(z), u - z \rangle - \varepsilon + \langle r, \bar{z} - z \rangle - \frac{1}{2} (\langle u - z, Q(u - z) \rangle - \langle u - \bar{z}, Q(u - \bar{z}) \rangle).\tag{2.20}$$

Primeiro note que esta relação com $u = z$ fornece

$$0 \leq \varepsilon - \langle r, \bar{z} - z \rangle - \langle z - \bar{z}, Q(z - \bar{z}) \rangle := \bar{\varepsilon}.\tag{2.21}$$

Por outro lado, usando a linearidade do produto interno e a simetria de Q , pode ser verificado que

$$\langle u - z, Q(u - z) \rangle - \langle u - \bar{z}, Q(u - \bar{z}) \rangle = -2 [\langle Q(z - \bar{z}), u - z \rangle - \langle z - \bar{z}, Q(z - \bar{z}) \rangle]. \quad (2.22)$$

Portanto, usando (2.22) em (2.20), temos que

$$h(u) - h(z) \geq \langle -\nabla q(z) + \delta, u - z \rangle - \bar{\varepsilon},$$

com $\delta := Q(z - \bar{z})$, o que prova a primeira relação em (2.17).

Para completar a prova invocamos (2.21) para ver que

$$\begin{aligned} 0 \leq \bar{\varepsilon} &= \varepsilon - \langle r, \bar{z} - z \rangle - \frac{1}{2} \langle z - \bar{z}, Q(z - \bar{z}) \rangle = \\ &= \varepsilon + \frac{1}{2} \langle r, Q^{-1}r \rangle - \left(\frac{1}{2} \langle r, Q^{-1}r \rangle + \langle Q^{-\frac{1}{2}}r, Q^{\frac{1}{2}}(\bar{z} - z) \rangle - \frac{1}{2} \langle z - \bar{z}, Q(z - \bar{z}) \rangle \right) = \\ &= \varepsilon + \frac{1}{2} \langle r, Q^{-1}r \rangle - \frac{1}{2} \|Q^{-\frac{1}{2}}r + Q^{\frac{1}{2}}(\bar{z} - z)\|^2, \end{aligned}$$

o que implica (2.17). ■

O seguinte resultado mostra que o Algoritmo 4 gera uma sequência de pontos que resolvem o Problema (P1).

Proposição 2.5 *Considere $\{\tilde{y}_k\}, \{\tilde{x}_k\}, \{r_x^k\}, \{r_y^k\}$ e $\{A_k\}$ gerados pelo Algoritmo 4 e a sequência $\{c_k\}$ dada por*

$$c_k = 1 + \frac{1}{A_k}, \quad \forall k \geq 1.$$

Então para todo $k \geq 1$ existe um número $\varepsilon_k \leq \frac{1}{2A_k} \|\tilde{x}_k - x_0\|^2$ tal que

$$\begin{aligned} r^k &:= \begin{pmatrix} r_x^k \\ r_y^k \end{pmatrix} \in \partial_{\varepsilon_k} [\hat{\Phi}(\cdot, \tilde{y}_k) - \hat{\Phi}(\tilde{x}_k, \cdot)](\tilde{x}_k, \tilde{y}_k) + \begin{pmatrix} c_k(\tilde{x}_k - x_0) \\ \tilde{y}_k - y_0 \end{pmatrix}, \\ \|r^k\|^2 + 2\varepsilon_k &\leq \hat{\sigma}_k \left(\|\tilde{x}_k - x_0\|^2 + \|\tilde{y}_k - y_0\|^2 \right), \end{aligned}$$

com

$$\hat{\sigma}_k := \frac{1}{A_k}.$$

Prova. Uma vez que o Algoritmo 4 é um caso especial do Algoritmo 1 para minimizar (2.14) com f_0 e g_0 dadas por (2.8) e (2.9), podemos usar (1.32) e obter

$$(f_0 + g_0)(\tilde{x}_k) := p_0(\tilde{x}_k) \leq (\Lambda_k + g_0)(x) + \frac{1}{2A_k} \|x - x_0\|^2, \quad \forall x \in X. \quad (2.23)$$

Agora, observe que pela definição de p_0 temos que para todo y vale

$$-p_0(\tilde{x}_k) \leq -\Phi(\tilde{x}_k, y) - g_0(\tilde{x}_k) + \frac{1}{2} \|y - y_0\|^2. \quad (2.24)$$

Sendo assim, aplicando o Lema 2.2 em (2.23), combinando com (2.24) e rearranjando os termos, temos que

$$\begin{aligned} & \Phi(x, \tilde{y}_k) + g_0(x) - \frac{1}{2} \|\tilde{y}_k - y_0\|^2 + \frac{1}{2A_k} \|x - x_0\|^2 - \\ & - \left[\Phi(\tilde{x}_k, y) + g_0(\tilde{x}_k) - \frac{1}{2} \|y - y_0\|^2 + \frac{1}{2A_k} \|\tilde{x}_k - x_0\|^2 \right] \geq -\frac{1}{2A_k} \|\tilde{x}_k - x_0\|^2, \end{aligned}$$

ou equivalentemente, definindo

$$\varepsilon'_k = \frac{1}{2A_k} \|\tilde{x}_k - x_0\|^2,$$

temos

$$0 \in \partial_{\varepsilon'_k} \left[\hat{\Phi}(\cdot, \tilde{y}_k) - \hat{\Phi}(\tilde{x}_k, \cdot) + \frac{c_k}{2} \|\cdot - x_0\|^2 + \frac{1}{2} \|\cdot - y_0\|^2 \right] (\tilde{x}_k, \tilde{y}_k).$$

Agora, aplicando a Proposição 2.4 na relação acima com $f = \hat{\Phi}(\cdot, \tilde{y}_k) - \hat{\Phi}(\tilde{x}_k, \cdot)$, $z = (\tilde{x}_k, \tilde{y}_k)$,

$q = \frac{1}{2} [\|\cdot - y_0\|^2 + c_k \|\cdot - x_0\|^2]$, temos que existe um vetor (\bar{x}, \bar{y}) satisfazendo

$$r' \in \partial_{\varepsilon_k} [\hat{\Phi}(\cdot, \tilde{y}_k) - \hat{\Phi}(\tilde{x}_k, \cdot)](\tilde{x}_k, \tilde{y}_k) + \begin{pmatrix} c_k(\tilde{x}_k - x_0) \\ \tilde{y}_k - y_0 \end{pmatrix}, \quad \text{com} \quad (2.25)$$

$$\varepsilon_k = \varepsilon'_k - \left[\frac{c_k}{2} \|\tilde{x}_k - \bar{x}\|^2 + \frac{1}{2} \|\tilde{y}_k - \bar{y}\|^2 \right],$$

$$r' = \begin{pmatrix} c_k(\bar{x} - \tilde{x}_k) \\ \bar{y} - \tilde{y}_k \end{pmatrix}. \quad (2.26)$$

Finalmente, por (2.25) e (2.26) temos que

$$\|r'\|^2 + 2\varepsilon_k = \varepsilon'_k \leq \frac{1}{A_k} [\|\tilde{x}_k - x_0\|^2 + \|\tilde{y}_k - y_0\|^2], \quad (2.27)$$

completando a demonstração. ■

Podemos verificar que esta última desigualdade mostra que o critério de parada no Algoritmo 4 garante que a condição de HPE (2.4) é satisfeita a cada passo com $\sigma(k) = 2/A_k$.

Com efeito, escrevendo

$$r_k := r' - \begin{pmatrix} c_k(\tilde{x}_k - x_0) \\ \tilde{y}_k - y_0 \end{pmatrix},$$

temos que vale (2.3). Vejamos que (2.27) implica (2.4):

$$\left\| r_k + \begin{pmatrix} \tilde{x}_k - x_0 \\ \tilde{y}_k - y_0 \end{pmatrix} \right\|^2 + 2\varepsilon_k =$$

$$\left\| r' - \frac{1}{A_k}(\tilde{x}_k - x_0) \right\|^2 + 2\varepsilon_k \leq \frac{2}{A_k} [\|\tilde{x}_k - x_0\|^2 + \|\tilde{y}_k - y_0\|^2],$$

esta última desigualdade segue de (2.27) e da desigualdade triangular.

3 ALGORITMO HPE PARA O PROBLEMA DE EQUILÍBRIO DE NASH

Neste capítulo apresentamos outra de nossas contribuições, que é a descrição de um algoritmo HPE para resolver o problema de equilíbrio de Nash. Na primeira seção descrevemos o problema de equilíbrio de Nash de duas variáveis, e na segunda especializamos o método HPE assim como fizemos no Capítulo 2 para o contexto do problema. Assim como no Capítulo 2 também apresentamos resultados de taxas de complexidade para a obtenção de uma solução aproximada usando o método.

3.1 O PROBLEMA DE EQUILÍBRIO DE NASH

Nesta seção descrevemos o problema de equilíbrio de Nash de duas variáveis.

O problema de equilíbrio de Nash tem diversas aplicações. Para ilustração consideramos a análise da teoria de jogos simultâneos, cujas estratégias e recompensas são de conhecimento comum a todos os jogadores. Resolver jogos simultâneos significa determinar quais serão suas ações e que consequências elas terão para os jogadores, desde que eles ajam racionalmente.

Informalmente, um conjunto de estratégias é um equilíbrio de Nash se nenhum jogador pode se dar melhor no jogo apenas mudando a sua estratégia. Imagine que cada jogador conhece as estratégias dos demais. Então suponha que cada jogador se pergunte se, conhecendo de antemão as estratégias dos demais, ele pode se beneficiar mudando sua estratégia. Se cada jogador após avaliação preferir não alterar sua estratégia então o conjunto de estratégias é um equilíbrio de Nash.

Mais formalmente, seja $(X, \hat{\Phi})$ um jogo de N jogadores onde X_i é a estratégia

do jogador i , $X = X_1 \times X_2 \times \cdots \times X_N$ é o conjunto de perfis de estratégia e $\hat{\Phi}(x) = (\hat{\Phi}_1(x), \dots, \hat{\Phi}_N(x))$ é a função de recompensa avaliada em $x \in X$. Seja x_i o perfil de estratégia do jogador i e x_{-i} o perfil de estratégia de todos os jogadores exceto o jogador i . Quando cada jogador $i \in \{1, \dots, N\}$ escolhe a estratégia x_i resultando no perfil $x = (x_1, \dots, x_N)$, então o jogador i recebe recompensa $\hat{\Phi}_i(x)$. Note que a recompensa depende do perfil de estratégia escolhido, ou seja, da estratégia escolhida do jogador i e também das estratégias escolhidas pelos demais. Um perfil de estratégia $x^* \in X$ é um equilíbrio de Nash se, conforme dito anteriormente, não há ganho para nenhum jogador i ao alterar sua estratégia, ou seja,

$$\hat{\Phi}_i(x_i^*, x_{-i}^*) \geq \hat{\Phi}_i(x_i, x_{-i}^*) \quad \forall i, x_i \in X_i.$$

Em nosso trabalho estamos interessados apenas em problemas de equilíbrio de Nash de duas variáveis, ou dois jogadores. Durante o restante deste capítulo, supomos que \mathcal{X} , \mathcal{Y} , \mathcal{Z} , X , Y , Z , $\langle \cdot, \cdot \rangle_{\mathcal{X}}$, $\langle \cdot, \cdot \rangle_{\mathcal{Y}}$, $\langle \cdot, \cdot \rangle$ são como na Subseção 1.3.1. Considere $\|\cdot\|_{\mathcal{X}}$, $\|\cdot\|_{\mathcal{Y}}$ e $\|\cdot\|$ normas induzidas associadas aos produtos internos de \mathcal{X} , \mathcal{Y} e \mathcal{Z} respectivamente.

A seguir enunciamos o problema de equilíbrio de Nash de duas variáveis na notação estabelecida acima. Dados $X \subseteq \mathcal{X}$ e $Y \subseteq \mathcal{Y}$ não-vazios, convexos e fechados, considere as funções $\hat{\Phi}_1 : X \times Y \rightarrow \mathbb{R}$ e $\hat{\Phi}_2 : X \times Y \rightarrow \mathbb{R}$. O problema de equilíbrio de Nash de duas variáveis determinado por $(\hat{\Phi}_1, \hat{\Phi}_2)$ e $Z = X \times Y$, denotado por $NE(\hat{\Phi}_1, \hat{\Phi}_2, Z)$, consiste em determinar um par $(x, y) \in X \times Y$ tal que

$$\hat{\Phi}_1(x, y) \leq \hat{\Phi}_1(x', y), \quad \hat{\Phi}_2(x, y) \leq \hat{\Phi}_2(x, y'), \quad \forall (x', y') \in X \times Y.$$

Supomos durante este capítulo que ambas funções $\hat{\Phi}_1$ e $\hat{\Phi}_2$ são da forma (2.2) e satisfazem B.1-B.3.

Assim como no caso do problema de ponto-de-sela, temos que se $(x, y) \in X \times Y$ e

$$(0, 0) \in \partial[\hat{\Phi}_1(\cdot, y) + \hat{\Phi}_2(x, \cdot)](x, y),$$

então (x, y) é uma solução de $NE(\hat{\Phi}_1, \hat{\Phi}_2, Z)$.

Considere $\varphi_1 : Y \rightarrow [-\infty, \infty)$ e $\varphi_2 : X \rightarrow [-\infty, \infty)$ dadas por

$$\varphi_1(y) := \inf_{x' \in X} \hat{\Phi}_1(x', y), \quad \varphi_2(x) := \inf_{y' \in Y} \hat{\Phi}_2(x, y'), \quad \forall (x, y) \in X \times Y. \quad (3.1)$$

Temos então que,

$$gap(x, y) := \hat{\Phi}_1(x, y) + \hat{\Phi}_2(x, y) - \varphi_1(y) - \varphi_2(x) \geq 0, \quad \forall (x, y) \in X \times Y.$$

Além disso, (x, y) é uma solução do problema de equilíbrio de Nash se e somente se $gap(x, y) = 0$, ou equivalentemente, (x, y) é uma solução com valor ótimo igual a zero para o problema de minimizar a função gap em $X \times Y$.

Agora introduzimos a seguinte definição de ponto de equilíbrio de Nash aproximado (isto é, uma solução aproximada do problema de equilíbrio de Nash $NE(\hat{\Phi}_1, \hat{\Phi}_2, Z)$).

Definição 3.1 Dados $(\rho, \varepsilon) \in \mathbb{R}_+ \times \mathbb{R}_+$, $z = (x, y) \in X \times Y$, $r \in \mathcal{X} \times \mathcal{Y}$ e $\tilde{\varepsilon} \in \mathbb{R}_+$, a tripla $(z, r, \tilde{\varepsilon})$ é dita um (ρ, ε) -ponto de equilíbrio de Nash de $NE(\hat{\Phi}_1, \hat{\Phi}_2, Z)$ se $\|r\| \leq \rho$, $\tilde{\varepsilon} \leq \varepsilon$ e

$$r \in \partial_{\tilde{\varepsilon}}[\hat{\Phi}_1(\cdot, y) + \hat{\Phi}_2(x, \cdot)](x, y). \quad (3.2)$$

Além disso, todo par $(r, \tilde{\varepsilon})$ é chamado um NE -resíduo para (x, y) com respeito à $NE(\hat{\Phi}_1, \hat{\Phi}_2, Z)$.

Vale a pena mencionar a relação entre um (ρ, ε) -ponto de equilíbrio de Nash e a noção de ε -ponto de equilíbrio de Nash. Dado $\varepsilon \geq 0$, chamamos um par $z = (x, y) \in X \times Y$ de ε -ponto de equilíbrio de Nash se satisfaz (3.2) com $(r, \tilde{\varepsilon}) = (0, \varepsilon)$. Para futura referência apresentamos a seguinte conhecida propriedade relacionando um ε -ponto de equilíbrio de Nash e um problema de inclusão monótona:

Lema 3.1 *Para todo $\varepsilon \geq 0$, temos que $\text{gap}(x, y) \leq \varepsilon$ se e somente se $z = (x, y)$ é um ε -ponto de equilíbrio de Nash, ou equivalentemente*

$$0 \in \partial_\varepsilon [\hat{\Phi}_1(\cdot, y) + \hat{\Phi}_2(x, \cdot)](x, y). \quad (3.3)$$

Prova. De fato, (3.3) ocorre se e somente se para todo $(x', y') \in X \times Y$ vale

$$\hat{\Phi}_1(x', y) + \hat{\Phi}_2(x, y') - [\hat{\Phi}_1(x, y) + \hat{\Phi}_2(x, y)] \geq \langle 0, (x' - x, y' - y) \rangle - \varepsilon = -\varepsilon,$$

o que, multiplicando a desigualdade por -1 e rearranjando os termos, é equivalente à $\text{gap}(x, y) \leq \varepsilon$. ■

Note que a condição acima para determinar se um par (x, y) é um ε -ponto de equilíbrio de Nash faz uso da hipótese de que as funções φ_1 e φ_2 dadas em (3.1) podem ser facilmente avaliadas em (x, y) . Entretanto, existem muitas aplicações onde isto não acontece. Por outro lado, supondo que $X \times Y$ é limitado, o seguinte resultado mostra que um algoritmo que gera uma seqüência de iterados $\{(z_k, r_k, \tilde{\varepsilon}_k)\}$ tais que $(z, r, \tilde{\varepsilon}) = (z_k, r_k, \tilde{\varepsilon}_k)$ satisfaz (3.2) e $\{(r_k, \tilde{\varepsilon}_k)\}$ converge a zero pode computar e detectar um ε -ponto de equilíbrio de Nash.

Lema 3.2 *Suponha que $X \times Y$ é limitado e seja D o seu diâmetro. Suponha também que $z = (x, y) \in X \times Y$, $r \in \mathcal{X} \times \mathcal{Y}$ e $\tilde{\varepsilon} \in \mathbb{R}_+$ são tais que $(z, r, \tilde{\varepsilon})$ satisfazem (3.2). Então, z é um ε' -ponto de equilíbrio de Nash de $NE(\hat{\Phi}_1, \hat{\Phi}_2, Z)$ onde*

$$\varepsilon' = \varepsilon'(z, r, \tilde{\varepsilon}) := \tilde{\varepsilon} + \max_{(x', y') \in X \times Y} \langle v, (x - x', y - y') \rangle. \leq \tilde{\varepsilon} + \|r\|D.$$

Prova. Segue imediatamente de (3.2) e da definição de ε -subdiferencial. ■

Portanto, sob a hipótese de que $X \times Y$ é limitado, segue do Lema 3.2 que se estamos em busca de um ε -ponto de equilíbrio de Nash, então podemos avaliar $\varepsilon_k := \varepsilon'(z_k, r_k, \tilde{\varepsilon}_k)$ e

terminar o algoritmo hipotético mencionado acima quando $\varepsilon_k \leq \varepsilon$. Note que a hipótese de limitação de $X \times Y$ garante que ε_k é finito para todo k . Além disso, uma vez que o algoritmo hipotético gera $\{(z_k, r_k, \tilde{\varepsilon}_k)\}$ de modo que $\{(r_k, \tilde{\varepsilon}_k)\}$ converge para zero, o algoritmo sempre obterá um ε -ponto de equilíbrio de Nash desta forma.

3.2 UMA VERSÃO DO MÉTODO HPE PARA RESOLVER O PROBLEMA DE EQUILÍBRIO DE NASH

A seguir estabelecemos, como apresentado no Capítulo 2, o método HPE para resolver o problema de inclusão monótona (3.2), e por consequência, o problema de equilíbrio de Nash $NE(\hat{\Phi}_1, \hat{\Phi}_2, Z)$.

Algoritmo 6: Método híbrido extragradiente proximal para resolver $NE(\hat{\Phi}_1, \hat{\Phi}_2, Z)$:

Passo 0 dados $(x_0, y_0) \in \mathcal{X} \times \mathcal{Y}$, $\lambda > 0$ e $0 \leq \sigma < 1$ faça $k = 1$;

Passo 1 encontre $(\tilde{x}_k, \tilde{y}_k) \in \mathcal{X} \times \mathcal{Y}$, $r_k = (r_k^x, r_k^y) \in \mathcal{X} \times \mathcal{Y}$ e $\varepsilon_k \geq 0$ tais que

$$(r_k^x, r_k^y) \in \partial_{\varepsilon_k} [\hat{\Phi}_1(\cdot, \tilde{y}_k) + \hat{\Phi}_2(\tilde{x}_k, \cdot)](\tilde{x}_k, \tilde{y}_k), \quad (3.4)$$

$$\begin{aligned} & \|\lambda r_k^x + \tilde{x}_k - x_{k-1}\|_{\mathcal{X}}^2 + \|\lambda r_k^y + \tilde{y}_k - y_{k-1}\|_{\mathcal{Y}}^2 + 2\lambda\varepsilon_k \\ & \leq \sigma^2 \left(\|\tilde{x}_k - x_{k-1}\|_{\mathcal{X}}^2 + \|\tilde{y}_k - y_{k-1}\|_{\mathcal{Y}}^2 \right); \end{aligned} \quad (3.5)$$

Passo 2 faça $x_k = x_{k-1} - \lambda r_k^x$, $y_k = y_{k-1} - \lambda r_k^y$ e $k \leftarrow k + 1$, e retorne ao Passo 1.

fim

A seguir, fazemos duas observações acerca do Algoritmo 6: Primeiramente, uma vez que o ε -subdiferencial está contido no ε -enlargamento (isto pode ser verificado de maneira análoga à feita na Proposição 1.8), o Algoritmo 6 é um caso especial do método HPE com

$\lambda_k := \lambda$. Na verdade, poderíamos tê-lo estabelecido em termos de uma seqüência variável de tamanhos de passo $\{\lambda_k\}$, mas fizemos $\lambda_k = \lambda$ por simplicidade. Em segundo lugar, similarmente ao método HPE, o Algoritmo 6 não especifica como determinar $(\tilde{x}_k, \tilde{y}_k)$, r_k e ε_k satisfazendo as condições (3.4) e (3.5).

O seguinte resultado descreve as propriedades de taxa de convergência pontual e ergódica do método HPE para o problema de equilíbrio de Nash.

Teorema 3.1 *Considere as seqüências $\{(\tilde{x}_k, \tilde{y}_k)\}$, $\{(r_k^x, r_k^y)\}$ e $\{\varepsilon_k\}$ geradas pelo Algoritmo 6 e defina para todo $k \in \mathbb{N}$,*

$$(\tilde{x}_k^a, \tilde{y}_k^a) := \frac{1}{k} \sum_{i=1}^k (\tilde{x}_i, \tilde{y}_i), \quad r_k^a := \frac{1}{k} \sum_{i=1}^k (r_i^x, r_i^y),$$

e

$$\varepsilon_k^a := \frac{1}{k} \sum_{i=1}^k [\varepsilon_i + \langle (\tilde{x}_i - \tilde{x}_k^a, \tilde{y}_i - \tilde{y}_k^a), (r_i^x, r_i^y) \rangle].$$

Seja d_0 a distância de (x_0, y_0) ao conjunto de soluções de $NE(\hat{\Phi}_1, \hat{\Phi}_2, Z)$. Então, para todo $k \in \mathbb{N}$, valem as seguintes afirmações:

- (a) *(taxa de convergência pontual) a tripla $((\tilde{x}_k, \tilde{y}_k), r_k, \varepsilon_k)$ é um $(\|r_k\|, \varepsilon_k)$ -ponto de equilíbrio de Nash de $NE(\hat{\Phi}_1, \hat{\Phi}_2, Z)$, ou equivalentemente (3.4) vale, e existe um índice $i \leq k$ tal que*

$$\|r_i\| \leq \frac{d_0}{\lambda} \sqrt{\frac{1+\sigma}{k(1-\sigma)}}, \quad \varepsilon_i \leq \frac{\sigma^2 d_0^2}{2k\lambda(1-\sigma^2)};$$

- (b) *(taxa de convergência ergódica) a tripla $((\tilde{x}_k^a, \tilde{y}_k^a), r_k^a, \varepsilon_k^a)$ é um $(\|r_k^a\|, \varepsilon_k^a)$ -ponto de equilíbrio de Nash de $NE(\hat{\Phi}_1, \hat{\Phi}_2, Z)$, ou equivalentemente*

$$r_k^a \in \partial_{\varepsilon_k^a}(\hat{\Phi}_1(\cdot, \tilde{y}_k^a) + \hat{\Phi}_2(\tilde{x}_k^a, \cdot))(\tilde{x}_k^a, \tilde{y}_k^a),$$

e

$$\|r_k^a\| \leq \frac{2d_0}{\lambda k}, \quad 0 \leq \varepsilon_k^a \leq \frac{2d_0^2}{\lambda k} \left(1 + \frac{\sigma}{\sqrt{1-\sigma^2}} \right).$$

Prova. A prova deste teorema é similar à do Teorema 3.4 de (HE; MONTEIRO, 2014). ■

3.3 UM ALGORITMO ACELERADO PARA O PROBLEMA DE EQUILÍBRIO DE NASH

Nesta seção estabelecemos um algoritmo acelerado para obter uma ε -solução do problema de equilíbrio de Nash usando a variante acelerada de Nesterov da Seção 1.4.

De modo análogo a feito no Capítulo 2, supomos a existência de funções próprias convexas e fechadas $\phi_i : \mathcal{X} \rightarrow (-\infty, +\infty]$, $i = 1, 2$ e funções $\Phi_i : \text{dom } \Phi_i \supset Z \rightarrow \mathbb{R}$ tais que

$$\text{dom } \phi_i = X,$$

$$\hat{\Phi}_i(x, y) := \Phi_i(x, y) + \phi_i(x), \quad \forall (x, y) \in Z, \quad i = 1, 2,$$

com Φ_i satisfazendo B.1-B.3. O problema é determinar um par (x, y) tal que

$$\begin{aligned} x &\in \operatorname{argmin}_{x \in X} \hat{\Phi}_1(x, y), \\ y &\in \operatorname{argmin}_{y \in Y} \hat{\Phi}_2(x, y). \end{aligned} \tag{3.6}$$

Com respeito à condição B.3, supomos que para todo $(x, y) \in X \times Y$ as funções $\Phi_1(\cdot, y)$ e $\Phi_2(\cdot, y)$ são diferenciáveis e existem constantes $L_{xx}^{(1)}$, $L_{xy}^{(1)}$, $L_{xx}^{(2)}$ e $L_{xy}^{(2)}$ satisfazendo

B.3') Para todos $(x, y), (x', y') \in X \times Y$ vale

$$\|\nabla_x \Phi_1(x', y') - \nabla_x \Phi_1(x, y)\|_{\mathcal{X}}^* \leq L_{xx}^{(1)} \|x - x'\|_{\mathcal{X}} + L_{xy}^{(1)} \|y - y'\|_{\mathcal{Y}},$$

$$\|\nabla_x \Phi_2(x', y') - \nabla_x \Phi_2(x, y)\|_{\mathcal{X}}^* \leq L_{xx}^{(2)} \|x - x'\|_{\mathcal{X}} + L_{xy}^{(2)} \|y - y'\|_{\mathcal{Y}}.$$

Adicionalmente, supomos ainda que para todo (x, y) as funções $\Phi_1(x, \cdot)$ e $\Phi_2(x, \cdot)$ são diferenciáveis e existem constantes $L_{yy}^{(1)}$, $L_{yx}^{(1)}$, $L_{yy}^{(2)}$ e $L_{yx}^{(2)}$ satisfazendo o análogo de B.3'

para a variável y . Considere as funções

$$\varphi_2(x) := \min_{y \in Y} \lambda \hat{\Phi}_2(x, y),$$

$$\varphi_1(y) := \min_{x \in X} \lambda \hat{\Phi}_1(x, y),$$

e a seguinte versão regularizada de (3.6):

$$\begin{aligned} x(y) &:= \operatorname{argmin}_{x \in X} \{ \lambda \hat{\Phi}_1(x, y) + \frac{1}{2} \|x - x_0\|^2 \}, \\ y(x) &:= \operatorname{argmin}_{y \in Y} \{ \lambda \hat{\Phi}_2(x, y) + \frac{1}{2} \|y - y_0\|^2 \}, \end{aligned} \quad (3.7)$$

para uma constante positiva $\lambda > 0$ que é o tamanho do passo e vetores x_0 e y_0 que serão os pontos iniciais. Defina ainda as funções

$$\begin{aligned} \varphi_2^0(x) &:= \min_{y \in Y} \lambda \hat{\Phi}_2(x, y) + \frac{1}{2} \|y - y_0\|^2, \\ \varphi_1^0(y) &:= \min_{x \in X} \lambda \hat{\Phi}_1(x, y) + \frac{1}{2} \|x - x_0\|^2, \\ g(x, y) &= \lambda [\phi_1(x) + \phi_2(y)], \quad g^0(x, y) = \lambda [\phi_1(x) + \phi_2(y)] + \frac{1}{2} \|x - x_0\|^2 + \frac{1}{2} \|y - y_0\|^2, \\ \hat{\Phi}_1^0(x, y) &:= \lambda \hat{\Phi}_1(x, y) + \frac{1}{2} \|x - x_0\|^2, \\ \hat{\Phi}_2^0(x, y) &:= \lambda \hat{\Phi}_2(x, y) + \frac{1}{2} \|y - y_0\|^2. \end{aligned} \quad (3.8)$$

Definimos então a função gap como

$$\operatorname{gap}(x, y) := \lambda [\hat{\Phi}_1^0(x, y) + \hat{\Phi}_2^0(x, y)] - \varphi_1^0(x) - \varphi_2^0(y) \geq 0. \quad (3.9)$$

A desigualdade é clara pela definição de φ_1^0 e φ_2^0 . Nosso objetivo será minimizar a função gap, ou seja, determinar o ponto (x, y) que faz valer $\varphi_1^0(x) = \varphi_2^0(y)$, ou equivalentemente determinar o minimizador de (3.9).

Assim como feito para o caso de ponto-de-sela, para simplicidade, consideramos a partir de agora $\lambda = 1$, de modo que o problema consiste em

$$\min_{(x, y)} [\Phi_1(x, y) + \Phi_2(x, y) - \varphi_2^0(x) - \varphi_1^0(y)] + \phi_1(x) + \phi_2(y) + \frac{1}{2} \|x - x_0\|^2 + \frac{1}{2} \|y - y_0\|^2. \quad (3.10)$$

A ideia central é usar a variante acelerada de Nesterov para otimização composta da Seção 1.4, onde a função entre colchetes é a parte “diferenciável” e o restante a parte convexa.

Chamaremos a função entre colchetes de

$$f_0(x, y) = (\Phi_1 + \Phi_2)(x, y) - \varphi_1^0(x) - \varphi_2^0(y), \quad (3.11)$$

e o restante de $g_0(x, y)$, isto é

$$g_0(x, y) = \phi_1(x) + \phi_2(y) + \frac{1}{2}\|x - x_0\|^2 + \|y - y_0\|^2. \quad (3.12)$$

O algoritmo é como a seguir:

Algoritmo 7: Algoritmo acelerado para resolver (3.10):

Passo 0 dados $\Phi_i, i = 1, 2, f_0, g_0$ como (3.11) e (3.12) e L^1 e L^2 as respectivas constantes de Lipschitz dos gradientes de Φ_1 e Φ_2 na norma de $X \times Y$,

faça

$$L := \left(L^1 + L^2 + 2 \left[L_{xx}^2 + \frac{1}{2}(L_{xy}^2)^2 \right] + 2 \left[L_{yy}^1 + \frac{1}{2}(L_{yx}^1)^2 \right] \right),$$

fixe $\sigma \in (0, 1)$, e uma função $\Lambda_0 \equiv 0$, sejam $z_- = (x_-, y_-) \in \mathcal{X} \times \mathcal{Y}, \sigma \in (0, 1)$, faça

$$A_0 = 0, \quad k = 1,$$

$$\hat{z}_0 = z_0 = P_{X \times Y}(\check{z}_0) := z_-;$$

Passo 1 faça

$$A_k = A_{k-1} + \frac{1 + \sqrt{1 + 4LA_{k-1}}}{2L},$$

$$\check{z}_k = \frac{A_{k-1}}{A_k} \check{z}_{k-1} + \frac{A_k - A_{k-1}}{A_k} z_{k-1},$$

compute a função afim Λ_k como

$$\Lambda_k := \frac{A_{k-1}}{A_k} \Lambda_{k-1} + \frac{A_k - A_{k-1}}{A_k} [f_0(\check{z}_k) + \langle \nabla f_0(\check{z}_k), \cdot - \check{z}_k \rangle_{\mathcal{X}}];$$

Passo 2 compute

$$z_k = \operatorname{argmin}_{z \in Z} \left\{ \Lambda_k(z) + g_0(z) + \frac{1}{2A_k} \|z - z_-\|^2 \right\},$$

$$\tilde{z}_k = \frac{A_{k-1}}{A_k} \tilde{z}_{k-1} + \frac{A_k - A_{k-1}}{A_k} z_k;$$

Passo 3 compute

$$z(\tilde{z}_k) = \begin{pmatrix} x(\tilde{y}_k) \\ y(\tilde{x}_k) \end{pmatrix}, \text{ com } x(\tilde{y}_k) \text{ e } y(\tilde{x}_k) \text{ dados por (3.7), } r_k = z_0 - z(\tilde{z}_k), \quad (3.13)$$

$$\varepsilon_k = \operatorname{gap}(\tilde{z}_k) = \hat{\Phi}_1^0(\tilde{x}_k, \tilde{y}_k) - \hat{\Phi}_1^0(x(\tilde{y}_k), \tilde{y}_k) + \hat{\Phi}_2^0(\tilde{z}_k) - \hat{\Phi}_2^0(\tilde{x}_k, y(\tilde{x}_k)); \quad (3.14)$$

Passo 4 se $\|r_k + \tilde{z}_k - z_0\|^2 + 2\varepsilon_k \leq \sigma^2 \|z_k - z_0\|^2$, termine. Caso contrário faça $k = k + 1$ e retorne ao Passo 1.

fim

É conveniente definir as funções

$$\hat{\Lambda}_k^0(z) := \Lambda_k(z) + g^0(z), \quad \hat{\Lambda}_k(z) := \Lambda_k(z) + g(z).$$

A partir de agora, para toda função $f : \mathcal{Z} \mapsto \mathbb{R}$ denotamos sua linearização em z_i como

$$l_f(z, z_i) = f(z_i) + \langle \nabla f(z_i), z - z_i \rangle,$$

e para todo índice fixado k denotamos

$$l_{f,k}(z) := \sum_{i=1}^k \frac{A_i - A_{i-1}}{A_k} l_f(z, z_i). \quad (3.15)$$

Observe que para quaisquer duas funções f_1 e f_2 temos que $l_{f_1,k} + l_{f_2,k} = l_{f_1+f_2,k}$. Usaremos este fato frequentemente sem fazer menção.

Observe ainda que, a fim de usar com sucesso a variante acelerada de Nesterov para o problema (3.9), as funções $\Lambda_k(z)$ e $g_0(z)$ devem satisfazer as hipóteses A.1-A.3. As hipóteses A.1-A.2 são claramente satisfeitas, basta mostrar que φ_i , $i = 1, 2$ satisfazem A.3. Sabemos que Φ_i são continuamente diferenciáveis. A fim de mostrar que φ_i , $i = 1, 2$ satisfazem A.3, podemos proceder de maneira similar à Proposição 2.1. Seguem as fórmulas das constantes de Lipschitz e dos gradientes.

Proposição 3.1 *As constantes $L_\lambda^{(1)}$ e $L_\lambda^{(2)}$ e as funções $\nabla f_1, \nabla f_2 : X \rightarrow X$ definidas, respectivamente, como*

$$L_\lambda^{(1)} := 2 \left(L_{xx}^{(1)} + (L_{xy}^{(1)})^2 \right), \quad L_\lambda^{(2)} := 2 \left(L_{xx}^{(2)} + (L_{xy}^{(2)})^2 \right),$$

$$\nabla f_1(x) := \nabla_x \Phi_1(x(y), y) \quad \nabla f_2(x) := \nabla_x \Phi_2(x, y(x)), \quad \forall x \in X,$$

onde $x(y)$ e $y(x)$ são definidas como

$$y(x) := \operatorname{argmax}_{y \in Y} \left\{ \Phi_2(x, y) - \frac{1}{2} \|y - y_-\|^2 \right\}, \quad \forall x \in X,$$

$$x(y) := \operatorname{argmax}_{x \in X} \left\{ \Phi_1(x, y) - \frac{1}{2} \|x - x_-\|^2 \right\}, \quad \forall y \in Y,$$

satisfazem a condição A.3 com $f = f_i$, $i = 1, 2$. Valem condições análogas para a variável y com as devidas modificações.

Prova. A prova é similar à da Proposição (2.1). ■

Segue da Proposição 3.1 que φ_1 e φ_2 satisfazem A.3 e os gradientes são dados por

$$\nabla \varphi_1^0(x) = \nabla_x \Phi_2(x, y(x)), \quad \nabla \varphi_1^0(y) = \nabla_y \Phi_1(x(y), y),$$

onde $y(x)$ e $x(y)$ são respectivamente os minimizadores nas definições de φ_1^0 e φ_2^0 em (3.8),

e as “constantes de Lipschitz” são dadas, respectivamente, por

$$L_{\varphi_1} := 2 \left[L_{xx}^{(2)} + (L_{xy}^{(2)})^2 \right], \quad L_{\varphi_2} := 2 \left[L_{yy}^{(1)} + (L_{yx}^{(1)})^2 \right],$$

onde usamos a notação de gradiente por facilidade de notação, tal como feito no Capítulo 2.

Este resultado implica que f_0 satisfaz A.3 com constante L dada pelo Algoritmo 7, e assim o método está bem definido.

3.4 RESOLVENDO O SUBPROBLEMA DE HPE

Nesta seção mostramos que \tilde{z}_k dado pelo Algoritmo 7 resolve o problema de ponto-de-sela aproximadamente, no sentido de que determinamos um vetor r_k e um escalar ε_k tais que

$$r_k \in \partial_{\varepsilon_k} \left[\hat{\Phi}_1(\cdot, \tilde{y}_k) + \hat{\Phi}_2(\tilde{x}_k, \cdot) \right] (\tilde{x}_k, \tilde{y}_k), \text{ com}$$

$$\|r_k + \tilde{z}_k - z_0\|^2 + 2\varepsilon_k \leq \sigma^2 \|\tilde{z}_k - z_0\|^2.$$

Na Subseção 3.4.1 provamos a convergência do método acelerado descrito como no Algoritmo 7. Observamos que este algoritmo requer a computação de $x(\tilde{y}_k)$ e $y(\tilde{x}_k)$ para verificar se o critério de parada é satisfeito ou não. Na Subseção 3.4.2 estabelecemos uma variante onde esta computação extra não é necessária, e também damos a prova de sua convergência.

3.4.1 Convergência do Algoritmo Acelerado Para o Problema de Equilíbrio de Nash

O seguinte lema técnico será utilizado nas demonstrações de nossos resultados de convergência desta subseção e também nos da Subseção 3.4.2.

Lema 3.3 *Seja $z^* = (x^*, y^*)$ o minimizador da função gap definida em (3.9). Então, valem*

$$\|z(\tilde{z}_k) - \tilde{z}_k\|^2 \leq \frac{1}{A_k} \|z^* - z_0\|^2 \quad (3.16)$$

e

$$\frac{1}{2}\|z^* - \tilde{z}_k\|^2 \leq \frac{1}{2A_k}\|z^* - z_0\|^2, \quad (3.17)$$

além disso, se $A_k > 2$ temos que

$$\|z^* - z_0\|^2 \leq \frac{2A_k}{A_k - 2}\|\tilde{z}_k - z_0\|^2. \quad (3.18)$$

Prova.

Primeiro, pela desigualdade (1.32) da Proposição 1.12 temos que para todo $z \in \mathcal{Z}$ vale

$$\frac{1}{2A_k}\|z - z_0\|^2 \geq \text{gap}(\tilde{z}_k) - \hat{\Lambda}_k^0(z) \geq \text{gap}(\tilde{z}_k) - \text{gap}(z) \geq \frac{1}{2}\|z - \tilde{z}_k\|^2 + \langle \nabla \text{gap}(z), \tilde{z}_k - z \rangle, \quad (3.19)$$

onde a última desigualdade deve-se ao fato de que $\hat{\Phi}^0$ é 1-fortemente convexa. Aqui, as funções gap e $\hat{\Lambda}_k^0$ fazem, respectivamente, os papéis de $f + g$ e $\Lambda_k + g$, e \tilde{z}_k e z_k os papéis de \tilde{x}_k e x_k . Uma vez que isto vale para todo z , em particular escolhendo $z = z^*$, vemos que vale (3.17), dado que $\nabla \text{gap}(z^*) = 0$.

Agora, observe que como $y(\tilde{x}_k)$ é por definição o minimizador da função 1-fortemente convexa $\hat{\Phi}_2^0(\tilde{x}_k, \cdot)$, e com isso $\nabla_y \hat{\Phi}_2^0(\tilde{x}_k, y(\tilde{x}_k)) = 0$, temos que

$$\|\tilde{y}_k - y(\tilde{x}_k)\|^2 = \|\tilde{y}_k - y(\tilde{x}_k)\|^2 + 2\langle \nabla_y \hat{\Phi}_2^0(\tilde{x}_k, y(\tilde{x}_k)), \tilde{y}_k - y(\tilde{x}_k) \rangle \leq 2 \left[\hat{\Phi}_2^0(\tilde{x}_k, \tilde{y}_k) - \hat{\Phi}_2^0(\tilde{x}_k, y(\tilde{x}_k)) \right]. \quad (3.20)$$

De modo análogo podemos chegar a

$$\|\tilde{x}_k - x(\tilde{y}_k)\|^2 \leq 2 \left[\hat{\Phi}_1^0(\tilde{x}_k, \tilde{y}_k) - \hat{\Phi}_1^0(x(\tilde{y}_k), \tilde{y}_k) \right]. \quad (3.21)$$

Somando (3.20) e (3.21), conseguimos que

$$\|\tilde{z}_k - z(\tilde{z}_k)\|^2 \leq 2\text{gap}(\tilde{z}_k) \leq 2\text{gap}(z^*) + \frac{1}{A_k}\|z^* - z_0\|^2 = \frac{1}{A_k}\|z^* - z_0\|^2,$$

a primeira desigualdade por (3.9), a segunda por causa de (1.32), e a igualdade porque sabemos que o gap vale 0 no seu minimizador, o que prova (3.16).

Agora, note que se k é suficientemente grande de modo que $A_k > 2$, temos por (3.17) que

$$\begin{aligned} \|z^* - z_0\|^2 &\leq 2 \left(\|z^* - \tilde{z}_k\|^2 + \|\tilde{z}_k - z_0\|^2 \right) \leq \frac{2}{A_k} \|z^* - z_0\|^2 + 2\|\tilde{z}_k - z_0\|^2 \Leftrightarrow \\ \|z^* - z_0\|^2 &\leq 2 \left(1 - \frac{2}{A_k} \right)^{-1} \|\tilde{z}_k - z_0\|^2 = \frac{2A_k}{A_k - 2} \|\tilde{z}_k - z_0\|^2, \end{aligned}$$

o que prova (3.18). ■

Proposição 3.2 *Considere $\{r_k\}$ e $\{\varepsilon_k\}$ gerados pelo Algoritmo 7, e z^* como no Lema 3.3. Então temos que para todo k ,*

$$r_k \in \partial_{\varepsilon_k} \left[\hat{\Phi}_1(\cdot, \tilde{y}_k) + \hat{\Phi}_2(\tilde{x}_k, \cdot) \right] (\tilde{x}_k, \tilde{y}_k), \quad (3.22)$$

com

$$\|r_k + \tilde{z}_k - z_0\|^2 + 2\varepsilon_k \leq \frac{2}{A_k} \|z^* - z_0\|^2. \quad (3.23)$$

Em particular, se k é suficientemente grande, vale

$$\|r_k + \tilde{z}_k - z_0\|^2 + 2\varepsilon_k \leq \frac{4}{A_k - 2} \|\tilde{z}_k - z_0\|^2. \quad (3.24)$$

Prova. Primeiro note que, como observado anteriormente, $z = (x, y)$ é um ε -ponto-de-equilíbrio de Nash de $NE(\Phi_1, \Phi_2, Z)$ se e somente se $\text{gap}(x, y) \leq \varepsilon$. Portanto, por (3.14), temos que

$$0 \in \partial_{\varepsilon_k} \left[\hat{\Phi}_1^0(\cdot, \tilde{y}_k) + \hat{\Phi}_2^0(\tilde{x}_k, \cdot) \right] (\tilde{x}_k, \tilde{y}_k).$$

Aplicando a Proposição 2.4 com $q = \frac{1}{2}\|z - z_0\|^2$, $h = \hat{\Phi}_1 + \hat{\Phi}_2$, $x = \hat{z}_k$, $\delta = \tilde{z}_k - z(\hat{z}_k)$, $\varepsilon = \varepsilon_k$, e comparando $\delta - \nabla q(\tilde{z}_k)$ com a definição de r_k dada por (3.13), temos que

$$r_k \in \partial_{\varepsilon'_k} \left[\hat{\Phi}_1(\cdot, \tilde{y}_k) + \hat{\Phi}_2(\tilde{x}_k, \cdot) \right] (\tilde{x}_k, \tilde{y}_k),$$

onde

$$\varepsilon'_k = \varepsilon_k - \frac{1}{2} \|\tilde{z}_k - z(\tilde{z}_k)\|^2 \geq 0,$$

o que claramente implica (3.22). Em particular esta desigualdade implica

$$\|\delta\|^2 = \|\tilde{z}_k - z(\tilde{z}_k)\|^2 \leq 2\varepsilon_k.$$

Agora esta estimativa fornece

$$\|r_k + \tilde{z}_k - z_0\|^2 + 2\varepsilon_k = \|\delta\|^2 + 2\varepsilon_k \leq 4\varepsilon_k. \quad (3.25)$$

Por outro lado, por (3.14) e pela desigualdade (3.19), sabemos que

$$\varepsilon_k = \text{gap}(\tilde{z}_k) - 0 = \text{gap}(\tilde{z}_k) - \text{gap}(z^*) \leq \frac{1}{2A_k} \|z^* - z_0\|^2,$$

o que combinado com (3.25) resulta em (3.23).

Finalmente, se k é suficientemente grande de modo que $A_k > 2$, podemos usar a relação (3.18) em (3.23), e obter (3.24). ■

3.4.2 Um Segundo Algoritmo Acelerado Para Resolver o Problema de Equilíbrio de Nash

O Algoritmo 7 faz o uso de uma minimização extra para obter $z(\tilde{z}_k)$. Nesta subseção apresentamos uma versão onde podemos nos livrar de tal minimização.

O algoritmo é o seguinte:

Algoritmo 8: Outro algoritmo acelerado para resolver (3.10):

Passo 0 considere $z_0 = (x_0, y_0)$ e sejam L^1 e L^2 as respectivas constantes de Lipschitz dos gradientes de Φ_1 e Φ_2 na norma de $X \times Y$;

Passo 1 repita exatamente os passos 2 e 3 do Algoritmo 7;

Passo 2 compute

$$c_k = 1 + \frac{1}{A_k}, \quad \tilde{r}_k = \begin{pmatrix} \frac{1}{A_k}(x_0 - x_k) \\ \frac{1}{A_k}(y_0 - y_k) \end{pmatrix},$$

$$\varepsilon_k = \frac{2}{A_k} \|\tilde{z}_k - z_0\|^2 - 2\langle \tilde{r}_k, x_k - \tilde{x}_k, y_k - \tilde{y}_k \rangle, \quad (3.26)$$

$$r_k = \begin{pmatrix} c_k(x_0 - x_k) + \sum_{i=1}^k \nabla_x \Phi_2(x_i, y(x_i)) - \nabla_x \Phi_2(\tilde{x}_k, \tilde{y}_k) \\ c_k(y_0 - y_k) + \sum_{i=1}^k \nabla_y \Phi_1(x(y_i), y_i) - \nabla_y \Phi_1(\tilde{x}_k, \tilde{y}_k) \end{pmatrix}; \quad (3.27)$$

Passo 3 se $\|r_k + \tilde{z}_k - z_0\|^2 + 2\varepsilon_k \leq \sigma^2 \|\tilde{z}_k - z_0\|^2$, termine, caso contrário faça $k = k + 1$ e retorne ao Passo 1.

fim

A seguir, apresentamos os resultados de convergência referentes ao Algoritmo 8.

Lema 3.4 *Defina*

$$\hat{\varepsilon}_k = \frac{1}{A_k} \|\tilde{z}_k - z_0\|^2 - \frac{1}{2A_k} \|\tilde{z}_k - z_k\|^2 + (\hat{\Lambda}_k^0 - gap)(\tilde{z}_k), \quad \tilde{r}_k = \frac{1}{A_k}(z_0 - z_k), \quad (3.28)$$

$$\varepsilon_r = \hat{\varepsilon}_k - \langle \tilde{r}_k, z_k - \tilde{z}_k \rangle - \frac{c_k}{2} \|\tilde{z}_k - z_k\|^2, \quad r'_k = c_k(z_0 - z_k), \quad \delta = c_k(\tilde{z}_k - z_k). \quad (3.29)$$

Então, temos que

$$r'_k \in \partial_{\varepsilon_r} \hat{\Lambda}_k(\tilde{z}_k), \quad (3.30)$$

com r'_k e ε_r satisfazendo

$$\varepsilon_r + \frac{1}{c_k} \|r'_k - \delta\|^2 = \hat{\varepsilon}_k + \frac{1}{2c_k} \|r'_k\|^2. \quad (3.31)$$

Prova. Usando o Lema 1.1 na desigualdade (1.32) da Proposição 1.12 com $f = gap$ e $\Lambda_k = \hat{\Lambda}_k^0$, temos que

$$\begin{aligned} \hat{\Lambda}_k^0(z) - gap(\tilde{z}_k) &\geq \frac{1}{2A_k} \left(\|z - z_k\|^2 - \|z - z_0\|^2 \right) = \\ &\left\langle \frac{1}{A_k}(z_0 - z_k), z - \tilde{z}_k \right\rangle + \frac{1}{2A_k} \left(\|\tilde{z}_k - z_k\|^2 - \|\tilde{z}_k - z_0\|^2 \right), \end{aligned} \quad (3.32)$$

onde na igualdade usamos que a função $\|\cdot - z_k\|^2 - \|\cdot - z_0\|^2$ coincide com sua expansão de Taylor de primeira ordem em torno de \tilde{z}_k . Observe que usando que $\hat{\Lambda}_k^0 \leq gap$ e esta última desigualdade com $z = \tilde{z}_k$, conseguimos

$$\|\tilde{z}_k - z_k\|^2 \leq \|\tilde{z}_k - z_0\|^2. \quad (3.33)$$

Notando que $\hat{\Lambda}_k^0 \leq \hat{\Lambda}_k^0 + \frac{1}{2A_k} \|\cdot - z_0\|^2$ e subtraindo $\frac{1}{2A_k} \|\tilde{z}_k - z_0\|^2$ de ambos os lados de (3.32) temos que

$$\begin{aligned} \hat{\Lambda}_k^0(z) + \frac{1}{2A_k} \|z - z_0\|^2 - \hat{\Lambda}_k^0(\tilde{z}_k) - \frac{1}{2A_k} \|\tilde{z}_k - z_0\|^2 &\geq \langle \tilde{r}_k, z - \tilde{z}_k \rangle + \\ \frac{1}{2A_k} \left(\|\tilde{z}_k - z_k\|^2 - \|\tilde{z}_k - z_0\|^2 \right) - \frac{1}{2A_k} \|\tilde{z}_k - z_0\|^2 &+ gap(\tilde{z}_k) - \hat{\Lambda}_k^0(\tilde{z}_k), \end{aligned}$$

ou equivalentemente,

$$\tilde{r}_k \in \partial_{\hat{\Lambda}_k^0 + \frac{1}{2A_k} \|\cdot - z_0\|^2}(\tilde{z}_k). \quad (3.34)$$

Observe que podemos invocar a Proposição 2.4 em (3.34) com $h = \hat{\Lambda}_k$, $q = \frac{c_k}{2} \|z - z_0\|^2$, $x = \tilde{z}_k$, e $\bar{x} = z_k = \operatorname{argmin}\{\hat{\Lambda}_k^0 + \frac{1}{2A_k} \|\cdot - z_0\|^2\}$. Tal proposição fornece $\delta = c_k(\tilde{z}_k - z_k)$ satisfazendo

$$-\nabla q(\tilde{z}_k) + \delta = c_k(z_0 - z_k) \in \partial_{\varepsilon_r} \hat{\Lambda}_k(\tilde{z}_k), \quad (3.35)$$

com ε_r dado por (3.29), o que prova (3.30) a qual, juntamente com a estimativa (2.17) fornece (3.31).

■

Lema 3.5 *Considere ε_k e r_k conforme (3.26) e (3.27), respectivamente. Então, temos que*

$$r_k \in \partial_{\varepsilon_k} [\hat{\Phi}_1(\cdot, \tilde{y}_k) + \hat{\Phi}_2(\tilde{x}_k, \cdot)](\tilde{x}_k, \tilde{y}_k). \quad (3.36)$$

Prova. A relação (3.35) fornece

$$\hat{\Lambda}_k(z) - \hat{\Lambda}_k(\tilde{z}_k) \geq \langle r'_k, z - \tilde{z}_k \rangle - \varepsilon_r. \quad (3.37)$$

Note que por (3.15) e convexidade de Φ_1 e Φ_2 , temos que $\hat{\Lambda}_k(z) \leq (\Phi_1 + \Phi_2)(z) - l_{\varphi_1^0, k}(y) - l_{\varphi_2^0, k}(x) + g(z)$. Usando isto e somando $l_{\Phi_1 + \Phi_2, k}(\tilde{z}_k) - (\Phi_1 + \Phi_2)(\tilde{z}_k)$ em ambos os lados de (3.37), temos que

$$\begin{aligned} & (\Phi_1 + \Phi_2 + g)(z) - (\Phi_1 + \Phi_2 + g)(\tilde{z}_k) - (l_{\varphi_1^0 + \varphi_2^0, k}(z) - l_{\varphi_1^0 + \varphi_2^0, k}(\tilde{z}_k)) \geq \\ & \langle r'_k, z - \tilde{z}_k \rangle - \varepsilon_r + [l_{\Phi_1 + \Phi_2, k} - (\Phi_1 + \Phi_2)](\tilde{z}_k). \end{aligned} \quad (3.38)$$

Em particular, fazendo $z = (x, \tilde{y}_k)$ na desigualdade (3.38) e escrevendo $r'_k = ((r'_k)^x, (r'_k)^y)$, temos que

$$\begin{aligned} & (\Phi_1 + \Phi_2)(x, \tilde{y}_k) + g_1(x) - (\Phi_1 + \Phi_2)(\tilde{x}_k, \tilde{y}_k) - g_1(\tilde{x}_k) - (l_{\varphi_2^0, k}(x) - l_{\varphi_2^0, k}(\tilde{x}_k)) \\ & \geq \langle (r'_k)^x, x - \tilde{x}_k \rangle - \varepsilon_r + [l_{\Phi_1 + \Phi_2, k} - (\Phi_1 + \Phi_2)](\tilde{z}_k) \Leftrightarrow \\ & \Phi_1(x, \tilde{y}_k) + g_1(x) - \Phi_1(\tilde{x}_k, \tilde{y}_k) - g_1(\tilde{x}_k) \geq \Phi_2(\tilde{x}_k, \tilde{y}_k) - \Phi_2(x, \tilde{y}_k) + l_{\varphi_2^0, k}(x) - l_{\varphi_2^0, k}(\tilde{x}_k) \\ & + \langle (r'_k)^x, x - \tilde{x}_k \rangle - \varepsilon_r + [l_{\Phi_1 + \Phi_2, k} - (\Phi_1 + \Phi_2)](\tilde{z}_k) \geq \\ & \langle (r'_k)^x + \nabla l_{\varphi_2^0, k}(\tilde{x}_k) - \nabla_x \Phi_2(\tilde{x}_k, \tilde{y}_k), x - \tilde{x}_k \rangle - \varepsilon_r + [l_{\Phi_1 + \Phi_2, k} - (\Phi_1 + \Phi_2)](\tilde{z}_k), \end{aligned} \quad (3.39)$$

onde na última desigualdade foi usada a convexidade de $l_{\varphi_2, k}(x)$ e $-\Phi_2(x, \hat{y}_k)$. Pelas definições de (3.29) e (3.28) temos que

$$\begin{aligned} \varepsilon_r + (\Phi_1 + \Phi_2 - l_{\Phi_1 + \Phi_2, k})(\tilde{z}_k) &= \frac{1}{A_k} \|\tilde{z}_k - z_0\|^2 - \frac{1}{2A_k} \|\tilde{z}_k - z_k\|^2 - \langle r_k, z_k - \tilde{z}_k \rangle - \frac{c_k}{2} \|\tilde{z}_k - z_k\|^2 \\ &\leq \frac{1}{A_k} \|\tilde{z}_k - z_0\|^2 - \langle r_k, z_k - \tilde{z}_k \rangle = \frac{\varepsilon_k}{2}. \end{aligned}$$

Substituindo esta relação em (3.39) conseguimos

$$\hat{\Phi}_1(x, \tilde{y}_k) - \hat{\Phi}_1(\tilde{x}_k, \tilde{y}_k) \geq \langle \nabla l_{\varphi_2^0, k}(\tilde{x}_k) - \nabla_x \Phi_2(\tilde{x}_k, \tilde{y}_k) + (r'_k)^x, x - \tilde{x}_k \rangle - \frac{\varepsilon_k}{2}. \quad (3.40)$$

Fazendo um processo similar para $z = (\tilde{x}_k, y)$ fixado na desigualdade (3.38) chegamos a

$$\hat{\Phi}_2(\tilde{x}_k, y) - \hat{\Phi}_2(\tilde{x}_k, \tilde{y}_k) \geq \langle \nabla l_{\varphi_1^0, k}(\tilde{y}_k) - \nabla_y \Phi_1(\tilde{x}_k, \tilde{y}_k) + (r'_k)^y, y - \tilde{y}_k \rangle - \frac{\varepsilon_k}{2}. \quad (3.41)$$

Por fim, somando (3.40) e (3.41) conseguimos a relação desejada (3.36). ■

Lema 3.6 *Considere z^* como no Lema 3.3, ε_k , r_k como obtido no Lema 3.5 e \tilde{z}_k conforme no Algoritmo 8. Então, temos que*

$$\|r_k + \tilde{z}_k - z_0\|^2 + 2\varepsilon_k \leq \sigma^* (\|\tilde{z}_k - z_0\| + \|z^* - z_0\|)^2, \quad (3.42)$$

e para k suficientemente grande, vale

$$\|r_k + \tilde{z}_k - z_0\|^2 + 2\varepsilon_k \leq \hat{\sigma} \|\tilde{z}_k - z_0\|^2, \quad (3.43)$$

com $\sigma^*, \hat{\sigma} = O(\frac{1}{\sqrt{A_k}})$.

Prova. Por (3.31), e pelas definições de $\hat{\varepsilon}_k$ e r_k , temos que

$$\frac{1}{c_k} \|\delta\|^2 \leq \hat{\varepsilon}_k + \frac{3}{2c_k} \|r_k\|^2 \leq \frac{1}{A_k} \|\tilde{z}_k - z_0\|^2 + \frac{3}{A_k} \|\tilde{z}_k - z_0\|^2 = \frac{4}{A_k} \|\tilde{z}_k - z_0\|^2, \quad (3.44)$$

onde na última desigualdade foi usado (3.33). Ainda, por (3.26) temos que

$$\varepsilon_k \leq \frac{1}{A_k} \|\tilde{z}_k - z_0\|^2 + \|r_k\| \|z_k - \tilde{z}_k\| \leq \frac{3}{A_k} \|\tilde{z}_k - z_0\|^2, \quad (3.45)$$

onde aqui novamente usamos (3.33).

Segue que

$$\begin{aligned}
\|r_k + \tilde{z}_k - z_0\|^2 + 2\varepsilon_k &= \left\| r'_k + \begin{pmatrix} \nabla l_{\varphi_2^0, k}(\tilde{x}_k) - \nabla_x \Phi_2(\tilde{x}_k, \tilde{y}_k) \\ \nabla l_{\varphi_1^0, k}(\tilde{y}_k) - \nabla_y \Phi_1(\tilde{x}_k, \tilde{y}_k) \end{pmatrix} + \tilde{z}_k - z_0 \right\|^2 + 2\varepsilon_k = \\
&= \left\| \frac{1}{A_k}(z_0 - \tilde{z}_k) + \delta + \begin{pmatrix} \nabla l_{\varphi_2^0, k}(\tilde{x}_k) - \nabla_x \Phi_2(\tilde{x}_k, \tilde{y}_k) \\ \nabla l_{\varphi_1^0, k}(\tilde{y}_k) - \nabla_y \Phi_1(\tilde{x}_k, \tilde{y}_k) \end{pmatrix} \right\|^2 + 2\varepsilon_k \\
&\leq \frac{2}{A_k} \|z_0 - \tilde{z}_k\|^2 + 2\|\delta\|^2 + 2\varepsilon_k + 2 \left\| \begin{pmatrix} \nabla l_{\varphi_2^0, k}(\tilde{x}_k) - \nabla_x \Phi_2(\tilde{x}_k, \tilde{y}_k) \\ \nabla l_{\varphi_1^0, k}(\tilde{y}_k) - \nabla_y \Phi_1(\tilde{x}_k, \tilde{y}_k) \end{pmatrix} \right\|^2 \Leftrightarrow \\
\|r_k + \tilde{z}_k - z_0\|^2 + 2\varepsilon_k &\leq \frac{8 + 8c_k}{A_k} \|\tilde{z}_k - z_0\|^2 + 2 \left\| \begin{pmatrix} \nabla l_{\varphi_2^0, k}(\tilde{x}_k) - \nabla_x \Phi_2(\tilde{x}_k, \tilde{y}_k) \\ \nabla l_{\varphi_1^0, k}(\tilde{y}_k) - \nabla_y \Phi_1(\tilde{x}_k, \tilde{y}_k) \end{pmatrix} \right\|^2.
\end{aligned} \tag{3.46}$$

As duas últimas desigualdades se devem a (3.44) e (3.45). Falta limitarmos o último termo.

Primeiro note que

$$\nabla l_{\varphi_2^0, k}(\tilde{x}_k) = \sum_{i=1}^k \frac{A_i - A_{i-1}}{A_k} \nabla l_{\varphi_2^0}(\tilde{x}_k, x_i) = \sum_{i=1}^k \frac{A_i - A_{i-1}}{A_k} \nabla_x \Phi_2(x_i, y(x_i)),$$

e similarmente para $\nabla l_{\varphi_1^0, k}(\tilde{y}_k)$. Segue então que

$$\begin{aligned}
\|\nabla_x \Phi_2(x_i, y(x_i)) - \nabla_x \Phi_2(\tilde{x}_k, \tilde{y}_k)\|^2 &= \|\nabla_x \Phi_2(x_i, y(x_i)) - \nabla_x \Phi_2(\tilde{x}_i, y(\tilde{x}_i)) + \nabla_x \Phi_2(\tilde{x}_i, y(\tilde{x}_i)) \\
&\quad - \nabla_x \Phi_2(\tilde{x}_i, \tilde{y}_i) + \nabla_x \Phi_2(\tilde{x}_i, \tilde{y}_i) - \nabla_x \Phi_2(x^*, y^*) + \nabla_x \Phi_2(x^*, y^*) - \nabla_x \Phi_2(\tilde{x}_k, \tilde{y}_k)\|^2 \\
&\leq 2\|\nabla \varphi_2^0(x_i) - \nabla \varphi_2^0(\tilde{x}_i)\|^2 + 2\|\nabla_x \Phi_2(\tilde{x}_i, y(\tilde{x}_i)) - \nabla_x \Phi_2(\tilde{x}_i, \tilde{y}_i)\|^2 + \\
&\quad 2\|\nabla_x \Phi_2(\tilde{x}_i, \tilde{y}_i) - \nabla_x \Phi_2(x^*, y^*)\|^2 + \|\nabla_x \Phi_2(x^*, y^*) - \nabla_x \Phi_2(\tilde{x}_k, \tilde{y}_k)\|^2 \leq \\
&\quad 2L_{\varphi_2}\|x_i - \tilde{x}_i\|^2 + 2L_{xy}^2\|y(\tilde{x}_i) - \tilde{y}_i\|^2 + 2(L_{xx}^2 + L_{xy}^2)\|\tilde{z}_i - z^*\|^2 + 2(L_{xx}^2 + L_{xy}^2)\|\tilde{z}_k - z^*\|^2.
\end{aligned}$$

Portanto, em vista do Lema 3.3 e das relações (1.32) e (1.31) da Proposição 1.12, (3.33) e

(3.17), conseguimos a estimativa

$$\begin{aligned}
& \left\| \sum_{i=1}^k \frac{A_i - A_{i-1}}{A_k} \nabla_x \Phi_2(x_i, y(x_i)) - \nabla_x \Phi_2(\tilde{x}_k, \tilde{y}_k) \right\|^2 \leq \\
& 2L_{\varphi_2} \sum_{i=1}^k \frac{A_i - A_{i-1}}{A_k} \|z_i - \tilde{z}_i\|^2 + 2L_{xy}^2 \sum_{i=1}^k \frac{A_i - A_{i-1}}{A_k} \|y(\tilde{x}_i) - \tilde{y}_i\|^2 + \\
& 2(L_{xx}^2 + L_{xy}^2) \sum_{i=1}^k \frac{A_i - A_{i-1}}{A_k} \|\tilde{z}_i - z^*\|^2 + 2(L_{xx}^2 + L_{xy}^2) \sum_{i=1}^k \frac{A_i - A_{i-1}}{A_k} \|\tilde{z}_k - z^*\|^2 \leq \\
& \frac{2L_{\varphi_2}}{LA_k} \|\tilde{z}_k - z_0\|^2 + \frac{2L_{xy}^2}{A_k} \sum_{i=1}^k \frac{A_i - A_{i-1}}{A_i} \|z^* - z_0\|^2 + \frac{2(L_{xy}^2 + L_{xx}^2)}{A_k} \sum_{i=1}^k \frac{A_i - A_{i-1}}{A_i} \|z^* - z_0\|^2 \\
& + \frac{2(L_{xy}^2 + L_{xx}^2)}{A_k} \sum_{i=1}^k \frac{A_i - A_{i-1}}{A_k} \|z^* - z_0\|^2 \leq \\
& \frac{2L_{\varphi_2}}{LA_k} \|\tilde{z}_k - z_0\|^2 + \left[\frac{2L_{xy}^2 k}{A_k} + \frac{2(L_{xy}^2 + L_{xx}^2)k}{A_k} \left(1 + \frac{1}{A_k}\right) \right] \|z^* - z_0\|^2.
\end{aligned}$$

Agora, lembrando que $\frac{k}{A_k} \leq \sqrt{\frac{4L}{A_k}}$ por causa de (1.30), temos que

$$\begin{aligned}
& \left\| \sum_{i=1}^k \frac{A_i - A_{i-1}}{A_k} \nabla_x \Phi_2(x(y_i), y_i) - \nabla_x \Phi_2(\tilde{x}_k, \tilde{y}_k) \right\|^2 \leq \\
& \frac{2L_{\varphi_2}}{LA_k} \|\tilde{z}_k - z_0\|^2 + \left[\frac{2L_{xy}^2 \sqrt{4L}}{A_k^{\frac{1}{2}}} + \frac{2(L_{xy}^2 + L_{xx}^2) \sqrt{4L}}{A_k^{\frac{1}{2}}} \left(1 + \frac{1}{A_k}\right) \right] \|z^* - z_0\|^2. \quad (3.47)
\end{aligned}$$

De modo similar, podemos chegar a

$$\begin{aligned}
& \left\| \sum_{i=1}^k \frac{A_i - A_{i-1}}{A_k} \nabla_y \Phi_1(x(y_i), y_i) - \nabla_y \Phi_1(\tilde{x}_k, \tilde{y}_k) \right\|^2 \leq \\
& \frac{2L_{\varphi_1}}{LA_k} \|\tilde{z}_k - z_0\|^2 + \left[\frac{2L_{yx}^1 \sqrt{4L}}{A_k^{\frac{1}{2}}} + \frac{2(L_{yx}^1 + L_{yy}^1) \sqrt{4L}}{A_k^{\frac{1}{2}}} \left(1 + \frac{1}{A_k}\right) \right] \|z^* - z_0\|^2. \quad (3.48)
\end{aligned}$$

Agora usando (3.47) e (3.48) em (3.46) obtemos

$$\begin{aligned}
& \|r_k + \tilde{z}_k - z_0\|^2 + 2\varepsilon_k \leq \left[\frac{8 + \frac{2(L_{\varphi_2} + L_{\varphi_1})}{L}}{A_k} \right] \|\tilde{z}_k - z_0\|^2 + \\
& \left[\frac{2(L_{xy}^2 + L_{yx}^1) \sqrt{4L}}{A_k^{\frac{1}{2}}} + \frac{2(L_{xy}^2 + L_{xx}^2 + L_{yx}^1 + L_{yy}^1) \sqrt{4L}}{A_k^{\frac{1}{2}}} \left(1 + \frac{1}{A_k}\right) \right] \|z^* - z_0\|^2, \quad (3.49)
\end{aligned}$$

o que prova (3.42).

Por fim, se k é suficientemente grande de modo que $A_k > 2$, podemos usar (3.18) em (3.49) e obter (3.43). ■

3.5 COMPLEXIDADE DOS ALGORITMOS ACELERADOS PARA O PROBLEMA DE EQUILÍBRIO DE NASH

Nesta seção apresentamos resultados sobre a complexidade dos algoritmos acelerados para o problema de equilíbrio de Nash apresentados na Seção 3.4.

Proposição 3.3 *O Algoritmo 7 termina após $\mathcal{O}(\lceil C\sqrt{L} \rceil)$ iterações para alguma constante C , enquanto o Algoritmo 8 após $\mathcal{O}(\lceil C(L)\sqrt{L} \rceil)$ onde $C(L)$ é da ordem de uma constante multiplicativa obtida em (3.49).*

Prova. Nos Lemas 3.2 e 3.6 respectivamente mostramos que se k é suficientemente grande de modo que $A_k > 2$, para o Algoritmo 7 existe $C > 0$ tal que

$$\|r_k + \tilde{z}_k - z_0\|^2 + 2\varepsilon_k \leq \frac{C}{A_k} \|\tilde{z}_k - z_0\|^2,$$

enquanto que para o Algoritmo 8 conseguimos uma constante $C(L)$ da ordem da constante obtida em (3.49) tal que

$$\|r_k + \tilde{z}_k - z_0\|^2 + 2\varepsilon_k \leq \frac{C(L)}{\sqrt{A_k}} \|\tilde{z}_k - z_0\|^2.$$

Agora, observe que em ambos os casos, se $k_0 \geq \sqrt{8L}$, temos que $A_k > 2$ para todo $k > k_0$.

De fato, pela relação (1.30), temos que

$$A_k > \frac{k_0^2}{4L} \geq 2.$$

Vamos mostrar que o Algoritmo 7 termina após $\mathcal{O}(\lceil k_0 = \max\{\sqrt{8L}, \sqrt{4LC}\sigma^{-2}\} \rceil)$ iterações.

Com efeito, suponha que o algoritmo não termina. Então, pelo critério de parada e pela definição de k_0 temos que

$$\begin{aligned} \sigma \|\tilde{z}_k - z_0\|^2 &< \|r_k + \tilde{z}_k - z_0\|^2 + 2\varepsilon_k \leq \frac{C}{A_k} \|\tilde{z}_k - z_0\|^2 \\ &\leq \frac{4LC}{k_0^2} \|\tilde{z}_k - z_0\|^2 \leq \sigma^2 \|\tilde{z}_k - z_0\|^2, \end{aligned}$$

o que é um absurdo. Segue que vale a afirmação, o que mostra que o Algoritmo 7 termina após $O(C\sqrt{L})$ iterações. De modo semelhante, podemos mostrar que escolhendo

$$k_0 = \max\{\sqrt{8L}, C(L)\sqrt{4L}\sigma^{-2}\},$$

o Algoritmo 8 termina após $O(\lceil C(L)\sqrt{L} \rceil)$ iterações. ■

Em todo caso, esta Proposição garante que a variante de Nesterov aplicada ao subproblema (3.9) demora $O(\lceil C(L)\sqrt{L} \rceil)$ iterações, onde $C(L)$ é a constante do caso do Algoritmo 7, e da mesma ordem da constante dada em (3.49) para o caso do Algoritmo 8.

Pode ser mostrado que com escolhas específicas de λ conseguimos uma complexidade ótima em certo sentido, mas não faremos tal análise aqui mas sim na posterior Seção 5.2, após derivarmos o algoritmo HPE para o caso não Euclidiano.

Em vista do resultado acima, podemos usar o Algoritmo 7 (ou 8) para resolver (3.9) aproximadamente conforme se segue.

Algoritmo 9: Algoritmo não-Euclidiano acelerado para o problema de equilíbrio de Nash:

Passo 0 dado $(x_0, y_0) \in \mathcal{X} \times \mathcal{Y}$, $\lambda > 0$ e $0 < \sigma < 1$, faça $k = 1$;

Passo 1 invoque o Algoritmo 7 com entrada

$$\hat{\Lambda}_k^0 = \lambda \hat{\Lambda}_k^0, \quad g^0 = \lambda g^0, \quad (u_0, v_0) = (x_{k-1}, y_{k-1}), \quad L = \lambda L,$$

e faça

$$(\tilde{x}_k, \tilde{y}_k) := (\tilde{u}, \tilde{v}), \quad r_k = (r_k^x, r_k^y) := \frac{1}{\lambda}(r^u, r^v), \quad \varepsilon_k := \frac{1}{\lambda}\tilde{\varepsilon},$$

onde (\tilde{u}, \tilde{v}) , (r^u, r^v) e $\tilde{\varepsilon}$ são gerados pela variante;

Passo 2 faça $x_k = x_{k-1} - \lambda r_k^x$, $y_k = y_{k-1} - \lambda r_k^y$, $k \leftarrow k + 1$, e retorne ao Passo 1.

fim

4 ALGORITMO NÃO-EUCLIDIANO HÍBRIDO EXTRAGRADIENTE PROXIMAL

Nos próximos dois capítulos apresentamos outra de nossas contribuições: o algoritmo não-Euclidiano híbrido extragradiante proximal para resolver problemas de ponto-de-sela. Neste capítulo descrevemos o método não-Euclidiano híbrido extragradiante proximal para resolver problemas de inclusão monótona e no capítulo seguinte iremos especializá-lo para o caso do problema de ponto-de-sela. Estabelecemos o método e suas propriedades de convergência na Seção 4.1 e apresentamos uma condição suficiente para obter a condição de erro na Seção 4.2.

Supomos durante esta seção que \mathcal{Z} é um espaço com produto interno denotado por $\langle \cdot, \cdot \rangle$, e que $\| \cdot \|$ é uma norma em \mathcal{Z} , não necessariamente induzida pelo produto interno.

4.1 MÉTODO NÃO-EUCLIDIANO HÍBRIDO EXTRAGRADIENTE PROXIMAL PARA PROBLEMAS DE INCLUSÃO MONÓTONA

Nesta subseção estabelecemos o método não-Euclidiano híbrido extragradiante proximal introduzido em (SOLODOV; SVAITER, 1999a) para determinar uma solução aproximada do problema de inclusão monótona (1.3) onde $T : \mathcal{Z} \rightrightarrows \mathcal{Z}$ é um operador monótono maximal tal que $\text{dom } T$ é limitado.

Antes de apresentar o método, precisamos das noções de funções geradoras de distância e distâncias de Bregman.

Definição 4.1 *Uma função própria, convexa e fechada $w : \mathcal{Z} \rightarrow [-\infty, \infty]$ é dita função geradora de distância se satisfaz as seguintes condições:*

- (i) $W := \text{dom } w$ é fechado e $W^0 := \text{int}(W) = \{z \in \mathcal{Z} : \partial w(z) \neq \emptyset\}$;

(ii) w restrita a W é contínua e é continuamente diferenciável em W^0 .

Ainda, a função $dw : \mathcal{Z} \times W^0 \rightarrow \mathbb{R}$ definida como

$$(dw)(z'; z) := w(z') - w(z) - \langle \nabla w(z), z' - z \rangle, \quad \forall z' \in \mathcal{Z}, \forall z \in W^0 \quad (4.1)$$

é referida como *distância de Bregman sobre W induzida por w* , ou simplesmente, uma *distância de Bregman sobre W quando w não precisa ser enfatizada*.

Por simplicidade, para todo $z \in W^0$, a função $(dw)(\cdot; z)$ será denotada por $(dw)_z$ de modo que

$$(dw)_z(z') = (dw)(z'; z), \quad \forall z \in W^0, \forall z' \in \mathcal{Z}.$$

As seguintes identidades seguem diretamente de (4.1):

$$\nabla(dw)_z(z') = -\nabla(dw)_{z'}(z) = \nabla w(z') - \nabla w(z), \quad \forall z, z' \in W^0, \quad (4.2)$$

$$(dw)_v(z') - (dw)_v(z) = \langle \nabla(dw)_v(z), z' - z \rangle + (dw)_z(z'), \quad \forall z' \in \mathcal{Z}, \forall v, z \in W^0. \quad (4.3)$$

O seguinte resultado clássico foi usado na Seção 1.4, mostrá-lo-emos agora, pois está estabelecida a base para sua compreensão.

Lema 4.1 *Considere um subconjunto convexo U de um espaço com produto interno \mathcal{Z} e $\psi, w : U \mapsto \mathbb{R}$ duas funções convexas e diferenciáveis, denote por dw a distância de Bregman gerada por w . Suponha que \hat{u} é a solução ótima de*

$$\min_{u \in U} \{\psi(u) + \eta dw_{u_0}(u)\},$$

para algum $u_0 \in U$ e $\eta > 0$. Então,

$$\psi(u) + \eta dw_{u_0}(u) \geq \min_{u \in U} \{\psi(u) + \eta dw_{u_0}(u)\} + \eta dw_{\hat{u}}(u), \quad \forall u \in U.$$

Prova. A definição de \hat{u} e o fato de que $\psi(\cdot) + \eta dw_{u_0}(\cdot)$ é uma função convexa diferenciável implicam

$$\langle \nabla \psi(\hat{u}) + \eta \nabla dw_{u_0}(\hat{u}), u - u_0 \rangle \geq 0, \quad \forall u \in U.$$

Usando (4.3), é fácil ver que

$$dw_{u_0}(u) = dw_{u_0}(\hat{u}) + \langle \nabla dw_{u_0}(\hat{u}), u - \hat{u} \rangle + dw_{\hat{u}}(u), \quad \forall u \in U.$$

Usando as duas relações acima e a hipótese de que ψ é convexa, concluímos, portanto que

$$\begin{aligned} \psi(u) + \eta dw_{u_0}(u) &= \psi(u) + \eta [dw_{u_0}(\hat{u}) + \langle \nabla dw_{u_0}(\hat{u}), u - \hat{u} \rangle + dw_{\hat{u}}(u)] \\ &\geq \psi(\hat{u}) + \eta dw_{u_0}(\hat{u}) + \langle \nabla \psi(\hat{u}) + \eta \nabla dw_{u_0}(\hat{u}), u - \hat{u} \rangle + \eta dw_{\hat{u}}(u) \\ &\geq \psi(\hat{u}) + \eta dw_{u_0}(\hat{u}) + \eta dw_{\hat{u}}(u), \end{aligned}$$

completando a prova. ■

A descrição do método não-Euclidiano híbrido extragradiante proximal é baseada em uma função geradora de distância w que também satisfaz a seguinte condição:

$$\text{C.1) } \text{dom } T \subset W.$$

De agora em diante, denotamos $Z := cl(\text{dom } T)$. Vemos, por C.1 e pela hipótese de W ser fechado, que $Z \subset W$.

A seguir formulamos o método não-Euclidiano híbrido extragradiante proximal.

Algoritmo 10: Método não-Euclidiano híbrido extragradiante proximal (NE-HPE):

Passo 0 tome $z_0 \in W^0$ e faça $j = 1$;

Passo 1 escolha $\sigma_j \in [0, 1]$, encontre $\lambda_j > 0$ e $(\tilde{z}_j, z_j, \varepsilon_j) \in W \times W^0 \times \mathbb{R}_+$ tais que

$$r_j := \frac{1}{\lambda_j} \nabla(dw)_{z_j}(z_{j-1}) \in T^{\varepsilon_j}(\tilde{z}_j), \quad (4.4)$$

$$(dw)_{z_j}(\tilde{z}_j) + \lambda_j \varepsilon_j \leq \sigma_j (dw)_{z_{j-1}}(\tilde{z}_j); \quad (4.5)$$

Passo 2 faça $j \leftarrow j + 1$ e vá retorne ao Passo 1.

fim

A seguir fazemos diversas observações sobre o Algoritmo 10. Primeiro, o Algoritmo 10 não especifica como determinar λ_j e $(\tilde{z}_j, z_j, \varepsilon_j)$ satisfazendo (4.4) e (4.5). O modo de determinar λ_j e $(\tilde{z}_j, z_j, \varepsilon_j)$ dependerá da versão do método em consideração e das propriedades do operador T . Em segundo lugar, se $\sigma_j = 0$, então (4.5) implica $\varepsilon_j = 0$ e $z_j = \tilde{z}_j$, e portanto $r_j \in T(z_j)$ por causa de (4.4). Segue que as condições (4.4)-(4.5) podem ser vistas como um relaxamento de uma iteração do método proximal não-Euclidiano exato, a saber,

$$\frac{1}{\lambda_j} \nabla(dw)_{z_j}(z_{j-1}) \in T(z_j).$$

Em terceiro lugar, se dw satisfaz algumas condições adicionais (i.e., a condição da Definição 4.2 que será apresentada posteriormente), então podemos mostrar (e será mostrado na Proposição 4.2) que a inclusão acima tem uma única solução z_j , disto concluimos que, para todo $\lambda_j > 0$, sempre é possível obter uma tripla $(\tilde{z}_j, z_j, \varepsilon_j)$ da forma $(z_j, z_j, 0)$ satisfazendo (4.4)-(4.5) com $\sigma_j = 0$. É claro, computar a tripla desta forma exata é computacionalmente caro, de modo que triplas inexatas $(\tilde{z}_j, z_j, \varepsilon_j)$ satisfazendo as condições com $\sigma_j > 0$ são computacionalmente mais atraentes.

É possível mostrar que alguns métodos conhecidos como aqueles em (KORPELEVIČ, 1976; TSENG, 2000) podem ser vistos como casos especiais do Algoritmo 10. No Capítulo

5 será apresentada uma versão do método acima para o contexto do problema de ponto-de-sela (onde $T = T_{\hat{\Phi}}$) no qual λ_j pode ser escolhido em um intervalo da forma $[\tau\lambda, \lambda]$ para $\lambda > 0$ fixado e $\tau \in (0, 1)$ e as triplas no Passo 1 são obtidas usando o algoritmo de gradiente acelerado da Seção 1.4 aplicado a um certo problema de otimização.

Na parte restante desta subseção, nos focamos em estabelecer complexidade ergódica e limites superiores para o número de iterações do Algoritmo 10. Começamos com alguns resultados preliminares.

Lema 4.2 *Para todo $j \geq 1$, são verdadeiras as seguintes afirmações:*

(a) *para todo $z \in W$, temos que*

$$(dw)_{z_{j-1}}(z) - (dw)_{z_j}(z) = (dw)_{z_{j-1}}(\tilde{z}_j) - (dw)_{z_j}(\tilde{z}_j) + \lambda_j \langle r_j, \tilde{z}_j - z \rangle;$$

(b) *para todo $z \in W$, temos que*

$$(dw)_{z_{j-1}}(z) - (dw)_{z_j}(z) \geq (1 - \sigma_j)(dw)_{z_{j-1}}(\tilde{z}_j) + \lambda_j (\langle r_j, \tilde{z}_j - z \rangle + \varepsilon_j).$$

Prova. (a) Usando (4.3) duas vezes e usando a definição de r_j dado em (4.4), temos que

$$\begin{aligned} (dw)_{z_{j-1}}(z) - (dw)_{z_j}(z) &= (dw)_{z_{j-1}}(z_j) + \langle \nabla(dw)_{z_{j-1}}(z_j), z - z_j \rangle \\ &= (dw)_{z_{j-1}}(z_j) + \langle \nabla(dw)_{z_{j-1}}(z_j), \tilde{z}_j - z_j \rangle + \langle \nabla(dw)_{z_{j-1}}(z_j), z - \tilde{z}_j \rangle \\ &= (dw)_{z_{j-1}}(\tilde{z}_j) - (dw)_{z_j}(\tilde{z}_j) + \langle \nabla(dw)_{z_{j-1}}(z_j), z - \tilde{z}_j \rangle \\ &= (dw)_{z_{j-1}}(\tilde{z}_j) - (dw)_{z_j}(\tilde{z}_j) + \lambda_j \langle r_j, \tilde{z}_j - z \rangle. \end{aligned}$$

(b) Esta afirmação segue como consequência imediata de (a) e de (4.5). ■

O seguinte resultado é uma consequência imediata do Lema 4.2 (b).

Lema 4.3 Para todo $j \geq 1$ e $z \in W$, temos que

$$(dw)_{z_0}(z) - (dw)_{z_j}(z) \geq \sum_{i=1}^j (1 - \sigma_i)(dw)_{z_{i-1}}(\tilde{z}_i) + \sum_{i=1}^j \lambda_i [\varepsilon_i + \langle r_i, \tilde{z}_i - z \rangle].$$

Prova. O lema segue somando a desigualdade no Lema 4.2 (b) de 1 a j . ■

Lema 4.4 Para todo $j \geq 1$, defina $\Lambda_j := \sum_{i=1}^j \lambda_i$,

$$\tilde{z}_j^a := \frac{1}{\Lambda_j} \sum_{i=1}^j \lambda_i \tilde{z}_i, \quad r_j^a := \frac{1}{\Lambda_j} \sum_{i=1}^j \lambda_i r_i, \quad \varepsilon_j^a := \frac{1}{\Lambda_j} \sum_{i=1}^j \lambda_i [\varepsilon_i + \langle r_i, \tilde{z}_i - \tilde{z}_j^a \rangle].$$

Então, temos que

$$\varepsilon_j^a \geq 0, \quad r_j^a \in T^{\varepsilon_j^a}(\tilde{z}_j^a), \quad (4.6)$$

$$\varepsilon_j^a + \langle r_j^a, \tilde{z}_j^a - z \rangle \leq \frac{(dw)_{z_0}(z)}{\Lambda_j}, \quad \forall z \in Z. \quad (4.7)$$

Prova. As relações em (4.6) seguem de (4.4) e da Proposição 1.10 (a). o Lema 4.3 e a hipótese de que $\sigma_j \in [0, 1]$, junto ao fato de que $Z \subset W$, e as definições ε_j^a e r_j^a , implicam que para todo $z \in Z$,

$$\begin{aligned} (dw)_{z_0}(z) &\geq (dw)_{z_0}(z) - (dw)_{z_j}(z) \geq \sum_{i=1}^j \lambda_i [\varepsilon_i + \langle r_i, \tilde{z}_i - z \rangle] = \\ &\sum_{i=1}^j \lambda_i [\varepsilon_i + \langle r_i, \tilde{z}_i - \tilde{z}_j^a \rangle + \langle r_i, \tilde{z}_j^a - z \rangle] = \Lambda_j [\varepsilon_j^a + \langle r_j^a, \tilde{z}_j^a - z \rangle], \end{aligned}$$

e portanto vale (4.7). ■

A fim de enunciar o resultado principal desta subseção, que estabelece a complexidade ergódica para o Algoritmo 10, definimos a grandeza

$$R = R(z_0; Z) := \sup \{(dw)_{z_0}(z) : z \in Z\} < \infty. \quad (4.8)$$

Observe que R é finito pela hipótese de que $\text{dom } T$ é limitado e pelos fatos de que $Z \subset W$ e que $(dw)_{z_0}(\cdot)$ é uma função contínua em W (pela Definição 4.1(ii)).

Teorema 4.1 Para todo $j \geq 1$, defina Λ_j , \tilde{z}_j^a , r_j^a e ε_j^a como no Lema 4.4 e também

$$\tilde{\varepsilon}_j := \varepsilon_j^a + \max \left\{ \langle r_j^a, \tilde{z}_j^a - z \rangle : z \in Z \right\}. \quad (4.9)$$

Então, valem as seguintes afirmações:

(a) para todo $j \geq 1$, temos que

$$0 \in T^{\tilde{\varepsilon}_j}(\tilde{z}_j^a), \quad \tilde{\varepsilon}_j \leq \frac{R}{\Lambda_j}; \quad (4.10)$$

(b) se $T = T_{\hat{\Phi}}$ onde $\hat{\Phi}$ é a função tal que para todo $z \in Z$ sua função estendida $\hat{\Phi}_z$ dada por (1.14) é convexa e fechada, e além disso satisfazendo a inclusão

$$r_j \in \partial_{\varepsilon_j}(\hat{\Phi}_{\tilde{z}_j})(\tilde{z}_j), \quad (4.11)$$

para todo $j \geq 1$, então temos que

$$0 \in \partial_{\tilde{\varepsilon}_j}(\hat{\Phi}_{\tilde{z}_j^a})(\tilde{z}_j^a), \quad \forall j \geq 1.$$

Prova. (a) A desigualdade (4.7), a definição de R dada em (4.8) e a definição de $\tilde{\varepsilon}_j$ dada em (4.9) implicam a desigualdade em (4.10). Defina $\delta_j := \tilde{\varepsilon}_j - \varepsilon_j^a$, note que (4.9) e as definições de ε -subdiferencial e de função indicadora dadas em (1.9) e (1.10) respectivamente, implicam $-r_j^a \in \partial_{\delta_j}(\mathcal{I}_Z)(\tilde{z}_j^a)$. Esta inclusão, juntamente com a inclusão em (4.6) e as Proposições 1.2 (a) e (1.7), implicam

$$\begin{aligned} 0 &\in T^{\varepsilon_j^a}(\tilde{z}_j^a) + \partial_{\delta_j}\mathcal{I}_Z(\tilde{z}_j^a) \subset T^{\varepsilon_j^a}(\tilde{z}_j^a) + (\partial\mathcal{I}_Z)^{\delta_j}(\tilde{z}_j^a) \subset (T + \partial\mathcal{I}_Z)^{\varepsilon_j^a + \delta_j}(\tilde{z}_j^a) \\ &= T^{\varepsilon_j^a + \delta_j}(\tilde{z}_j^a) = T^{\tilde{\varepsilon}_j}(\tilde{z}_j^a), \end{aligned}$$

onde a última desigualdade deve-se pela definição de δ_j .

(b) Esta afirmação pode ser demonstrada usando um argumento similar aos utilizados nas inclusões do Lema 4.4 e no item (a), exceto que invocamos a Proposição 1.10 (b) ao invés da Proposição 1.10 (a). ■

Note que podemos calcular $\tilde{\varepsilon}_j$ em (4.9) para as instâncias em que a minimização de uma função linear em Z pode ser facilmente computada. Note ainda que se Λ_j tende a ∞ , a relação (4.10) implica que todo ponto limite de \tilde{z}_j^a é uma solução de (1.3). A desigualdade nesta relação estabelece uma taxa de convergência de \tilde{z}_j^a , medida em termos do tamanho de $\tilde{\varepsilon}_j$, que é da ordem de $\mathcal{O}(1/\Lambda_j)$. Claramente, esta taxa de convergência se reduz a $\mathcal{O}(1/j)$ para o caso em que a sequência de tamanhos do passo $\{\lambda_j\}$ é constante.

4.2 UMA CONDIÇÃO SUFICIENTE PARA AS CONDIÇÕES DE ERRO

Nesta subseção, discutimos como obter uma condição suficiente que assegura as condições de erro (4.4)-(4.5) para classes especiais de operadores monótonos maximais e distâncias de Bregman.

Observe que dados $z_- = z_{j-1} \in W^0$ e $\sigma = \sigma_j \in [0, 1]$, cada iteração do Algoritmo 10 envolve a computação de um tamanho de passo $\lambda = \lambda_j > 0$ e uma tripla $(\tilde{z}, z, \varepsilon) = (\tilde{z}_j, z_j, \varepsilon_j) \in W \times W^0 \times \mathbb{R}_+$ satisfazendo as condições de erro:

$$\frac{1}{\lambda} \nabla(dw)_z(z_-) \in T^\varepsilon(\tilde{z}), \quad (4.12)$$

$$(dw)_z(\tilde{z}) + \lambda\varepsilon \leq \sigma(dw)_{z_-}(\tilde{z}). \quad (4.13)$$

Nosso objetivo nesta subseção é determinar uma condição suficiente para obter tal tripla. Nossa discussão nesta subseção se aplica a operadores monótonos maximais satisfazendo a seguinte condição:

D.1) para todo $\tilde{z} \in \text{dom } T$, existe uma função própria, convexa e fechada $f_{\tilde{z}} : \mathcal{Z} \rightarrow [-\infty, \infty]$ tal que $\tilde{z} \in \text{dom}(f_{\tilde{z}})$ e $\partial_\varepsilon(f_{\tilde{z}})(\tilde{z}) \subset T^\varepsilon(\tilde{z})$ para todo $\varepsilon \geq 0$.

Note que a Proposição 1.8 implica que o operador $T_{\hat{\Phi}}$ definido em (1.15) satisfaz a condição D.1 com $f_{\tilde{z}} = \hat{\Phi}_{\tilde{z}}$ para todo $\tilde{z} \in \text{dom } T_{\hat{\Phi}}$, onde $\hat{\Phi}_{\tilde{z}}$ é dada como em (1.14).

A fim de enunciar o resultado principal desta subseção, necessitamos introduzir uma classe especial de distâncias de Bregman que também será utilizada em outras partes deste trabalho.

Definição 4.2 *Dado um escalar $\mu > 0$ e um conjunto convexo não-vazio $\Omega \subset \mathcal{Z}$, uma distância de Bregman dw sobre W é chamada de (μ, Ω) -distância de Bregman sobre W se $\Omega \cap W^0 \neq \emptyset$ e*

$$(dw)_z(z') \geq \frac{\mu}{2} \|z' - z\|^2, \quad \forall z, z' \in \Omega \cap W^0. \quad (4.14)$$

A seguir, fazemos algumas observações acerca da definição acima. Primeiramente, para todo $z \in \Omega \cap W^0$, a desigualdade em (4.14) vale para todo $z' \in \Omega \cap W$ pela continuidade de $(dw)_z(\cdot)$. Em segundo lugar, (4.14) é equivalente à função geradora de distância w ser μ -fortemente convexa em $\Omega \cap W$.

Antes de enunciar o resultado principal desta subseção precisamos de dois resultados de existência e unicidade de soluções de certos problemas regularizados de otimização convexa e de inclusão monótona. Primeiro enunciamos sem demonstração um resultado conhecido a respeito de problemas de otimização convexa regularizados. Tal resultado não precisa ser mostrado, pois é consequência do próximo resultado demonstrado na sequência.

Proposição 4.1 *Considere uma função própria, convexa e fechada $\psi : \mathcal{Z} \rightarrow [-\infty, \infty]$ e, para algum $\mu > 0$, suponha que dw é uma $(\mu, \text{dom } \psi)$ -distância de Bregman sobre W conforme a Definição 4.2. Então, para todo $z_- \in W^0 \cap \text{dom } \psi$, o problema*

$$\inf \{ \psi(u) + (dw)_{z_-}(u) : u \in \mathcal{Z} \}$$

tem uma única solução global z , que necessariamente está em $W^0 \cap \text{dom } \psi$. Ainda, z é o único zero da inclusão $\nabla w(z_-) \in (\partial\psi + \partial w)(z)$.

O próximo resultado generaliza a Proposição 4.1 para o contexto de problemas regularizados de inclusão monótona.

Proposição 4.2 *Considere um operador monótono maximal $T : \mathcal{Z} \rightrightarrows \mathcal{Z}$ e, para algum $\mu > 0$, suponha que dw é uma $(\mu, \text{dom } T)$ -distância de Bregman sobre W conforme a Definição 4.2. Então, para todo $z' \in \mathcal{Z}$, a inclusão*

$$z' \in (T + \partial w)(z)$$

tem uma única solução z (que necessariamente pertence a $W^0 \cap \text{dom } T$ pela Definição 4.1(i)).

Prova. Defina $\bar{w} := w + \mathcal{I}_Z$ onde $Z := \text{cl}(\text{dom } T)$. Afirmamos que \bar{w} é uma função própria fechada e μ -fortemente convexa tal que

$$\text{dom } \partial \bar{w} = Z \cap W^0, \quad T + \partial w = T + \partial \bar{w}.$$

De fato, pela Proposição 1.1, o conjunto Z é convexo e fechado, e assim a função indicadora \mathcal{I}_Z também é convexa e fechada. Esta conclusão junto com a Definição 4.2 implicam que \bar{w} é uma função própria, fechada e μ -fortemente convexa. Uma vez que a hipótese de que dw é uma $(\mu, \text{dom } T)$ -distância de Bregman sobre W implica que $W^0 \cap \text{dom } T \neq \emptyset$ e W^0 é aberto, vemos que $W^0 \cap \text{ri } Z = W^0 \cap \text{ri}(\text{dom } T) \neq \emptyset$, e portanto que os interiores relativos dos domínios das funções convexas w e \mathcal{I}_Z se intersectam. Esta conclusão junto com a Proposição 1.4 implicam $\partial \bar{w} = \partial(w + \mathcal{I}_Z) = \partial w + \partial \mathcal{I}_Z$, e portanto que $\text{dom } \partial \bar{w} = \text{dom } \partial \mathcal{I}_Z \cap \text{dom } \partial w = Z \cap W^0$. Agora, como T é maximal, com a Proposição 1.7 e a última conclusão, temos que

$$T + \partial \bar{w} = T + (\partial w + \partial \mathcal{I}_Z) = (T + \partial \mathcal{I}_Z) + \partial w = T + \partial w,$$

e portanto a afirmação está provada.

A seguir estabeleceremos a conclusão do lema. Alterando μ se necessário, podemos supor, sem perda de generalidade, que $\|\cdot\|$ é a norma associada ao produto interno $\langle \cdot, \cdot \rangle$. Uma vez que \bar{w} é uma função própria, fechada e μ -fortemente convexa, segue da afirmação anterior e da Proposição 1.11 que $\bar{w}_0 := \bar{w} - \mu\|\cdot\|^2/2$ é uma função própria, convexa e fechada. Agora, defina $\bar{T} := T + \partial\bar{w}_0$ e note que

$$T + \partial w = T + \partial\bar{w} = \bar{T} + \mu I,$$

pela afirmação acima e pelo fato de que $\partial\bar{w} = \partial\bar{w}_0 + \mu I$. Assim, a conclusão do Lema seguirá do Teorema de Minty (Teorema 1.1) se \bar{T} for monótono maximal. Com efeito, primeiro note que a afirmação acima implica

$$\text{ri}(\text{dom } \partial\bar{w}) = \text{ri}(Z \cap W^0) = \text{ri } Z \cap W^0,$$

pelo fato de que o último conjunto é não-vazio. Uma vez que tanto T quanto $\partial\bar{w}_0$ são monótonos maximais (o segundo pela Proposição 1.3) e a interseção dos interiores relativos de seus domínios é $\text{ri } Z \cap W^0 \neq \emptyset$, segue da Proposição 1.5 que \bar{T} é monótono maximal, completando a prova. ■

Agora apresentamos a condição suficiente anteriormente mencionada para se obter uma tripla satisfazendo (4.12)-(4.13).

Proposição 4.3 *Suponha que T é um operador monótono maximal satisfazendo a condição C.1 e que dw é uma $(\mu, \text{dom } T)$ -distância de Bregman sobre W para algum $\mu > 0$. Considere $\lambda > 0$, $z_- \in W^0$ e $\tilde{z} \in \text{dom } T \cap W^0$. Suponha que existe uma função própria e fechada tal que $\Gamma_{\tilde{z}} \leq f_{\tilde{z}}$ e defina as grandezas*

$$z := \text{argmin}_u \left\{ \Gamma_{\tilde{z}}(u) + \frac{1}{\lambda}(dw)_{z_-}(u) \right\} \quad e \quad (4.15)$$

$$\varepsilon := f_{\tilde{z}}(\tilde{z}) - \Gamma_{\tilde{z}}(z) - \langle r, \tilde{z} - z \rangle, \quad (4.16)$$

onde

$$r := \frac{1}{\lambda} \nabla(dw)_z(z_-). \quad (4.17)$$

Então, valem as seguintes afirmações:

(a) z está bem definido, $\varepsilon \in [0, \infty)$ e vale a seguinte inclusão (que implica (4.12) devido a D.1):

$$r \in \partial_\varepsilon(f_{\tilde{z}})(\tilde{z}); \quad (4.18)$$

(b) se, adicionalmente, para um dado escalar $\sigma \geq 0$ temos que

$$f_{\tilde{z}}(\tilde{z}) + \frac{1-\sigma}{\lambda}(dw)_{z_-}(\tilde{z}) \leq \inf \left\{ \Gamma_{\tilde{z}}(u) + \frac{1}{\lambda}(dw)_{z_-}(u) : u \in \mathcal{Z} \right\}, \quad (4.19)$$

então vale (4.13).

Prova. (a) As hipóteses de que $\tilde{z} \in \text{dom } T \cap W^0$ e $\Gamma_{\tilde{z}} \leq f_{\tilde{z}}$ junto com a condição D.1 implicam $\tilde{z} \in \text{dom } f_{\tilde{z}} \cap W^0 \subset \text{dom } \Gamma_{\tilde{z}} \cap W^0$. Uma vez que dw é uma $(\mu, \text{dom } T)$ -distância de Bregman sobre W , segue de (4.17) e da Proposição 4.1 com $\psi = \lambda\Gamma_{\tilde{z}}$ que z está bem definido e satisfaz

$$r \in \partial\Gamma_{\tilde{z}}(z). \quad (4.20)$$

Vemos que a última conclusão e o fato de que $\tilde{z} \in \text{dom } f_{\tilde{z}}$ implicam $\varepsilon < \infty$. Usando a hipótese de que $\Gamma_{\tilde{z}} \leq f_{\tilde{z}}$, e as relações (4.16) e (4.20), concluimos que

$$f_{\tilde{z}}(u) \geq \Gamma_{\tilde{z}}(u) \geq \Gamma_{\tilde{z}}(z) + \langle r, u - z \rangle = f_{\tilde{z}}(\tilde{z}) + \langle r, u - \tilde{z} \rangle - \varepsilon, \quad \forall u \in \mathcal{Z}, \quad (4.21)$$

e portanto a primeira inclusão em (4.18) é assegurada. Usando a desigualdade (4.21) com $u = \tilde{z}$ temos que $\varepsilon \geq 0$.

(b) Note que (4.19) e (4.15) implicam

$$f_{\tilde{z}}(\tilde{z}) + \frac{1-\sigma}{\lambda}(dw)_{z_-}(\tilde{z}) \leq \Gamma_{\tilde{z}}(z) + \frac{1}{\lambda}(dw)_{z_-}(z). \quad (4.22)$$

Além disso, as relações (4.2), (4.3) e (4.17) implicam

$$(dw)_{z-}(\tilde{z}) - (dw)_{z-}(z) = (dw)_z(\tilde{z}) + \langle \nabla(dw)_{z-}(z), \tilde{z} - z \rangle = (dw)_z(\tilde{z}) - \lambda \langle r, \tilde{z} - z \rangle. \quad (4.23)$$

Agora, usando (4.16), (4.22) e (4.23), concluimos que

$$\begin{aligned} (dw)_z(\tilde{z}) + \lambda \varepsilon &= (dw)_z(\tilde{z}) + \lambda [f_{\tilde{z}}(\tilde{z}) - \Gamma_{\tilde{z}}(z) - \langle r, \tilde{z} - z \rangle] \\ &= (dw)_{z-}(\tilde{z}) - (dw)_{z-}(z) + \lambda [f_{\tilde{z}}(\tilde{z}) - \Gamma_{\tilde{z}}(z)] \leq \sigma(dw)_{z-}(\tilde{z}), \end{aligned}$$

e portanto vale (4.13). ■

5 UMA VERSÃO ACELERADA DO MÉTODO NE-HPE

Neste capítulo apresentamos nossa contribuição mencionada anteriormente: uma versão acelerada do Algoritmo 10 para resolver o problema de ponto-de-sela no contexto de distâncias de Bregman conforme estabelecidas no capítulo anterior. Durante este capítulo, supomos que $\mathcal{X}, \mathcal{Y}, \mathcal{Z}, X, Y, Z, \langle \cdot, \cdot \rangle_{\mathcal{X}}, \langle \cdot, \cdot \rangle_{\mathcal{Y}}, \langle \cdot, \cdot \rangle$ e $\hat{\Phi}$ são como na Subseção 1.3.1. Além disso, denote por $\|\cdot\|_{\mathcal{X}}$ e $\|\cdot\|_{\mathcal{Y}}$ as normas dadas em \mathcal{X} e \mathcal{Y} , respectivamente, que não necessariamente são as induzidas pelos produtos internos dos respectivos espaços.

A seguir descrevemos a distância de Bregman utilizada em nossa versão. Considere dw^X (respectivamente, dw^Y) uma (η_X, X) -distância de Bregman (respectivamente, (η_Y, Y) -distância de Bregman) sobre $W_X \subset \mathcal{X}$ (respectivamente, $W_Y \subset \mathcal{Y}$). Fazendo $W = W_X \times W_Y$ e $W^0 = \text{int}(W)$, usando as propriedades de dw^X e dw^Y , podemos verificar que a função dw definida como

$$(dw)_z(z') := (dw^X)_x(x') + (dw^Y)_y(y'), \quad \forall z = (x, y) \in W^0, \forall z' = (x', y') \in W, \quad (5.1)$$

é uma distância de Bregman sobre W induzida por $w^X + w^Y$.

Supomos que $Z \subset W$ para assegurar que o operador $T = T_{\hat{\Phi}}$ dado por (1.15) satisfaz a condição C.1, e portanto os resultados da Seção 4.1 ainda valem no presente contexto.

A fim de descrever a nossa versão, será suficiente explicar como é implementado o Passo 1 do Algoritmo 10. Este será o tema da Seção 5.1 abaixo, na qual descrevemos um esquema para implementar este passo baseando-se no algoritmo de gradiente acelerado da Seção 1.4. Por ora, apenas mencionamos que o tamanho do passo λ_j não será escolhido constante mas computado dentro de um intervalo da forma $[\tau\lambda, \lambda]$ onde $\lambda > 0$ e $\tau \in (0, 1)$ são fixados em nossa versão. Adicionalmente, o esquema da Seção 5.1 também descreve

como computar a tripla $(\tilde{z}_j, z_j, \varepsilon_j)$ satisfazendo a condição (4.5) com dw dado por (5.1), e a inclusão mais forte (4.11).

Mais especificamente, na Seção 5.1 descrevemos um esquema para resolver o seguinte problema:

(P2) *Dado um par $z_- = (x_-, y_-) \in W^0$, e escalares $\sigma \in (0, 1]$, $\lambda > 0$ e $\tau \in (0, 1)$, o problema consiste em determinar $\tilde{\lambda} \in [\tau\lambda, \lambda]$ e uma tripla $(\tilde{z}, z, \varepsilon) \in W \times W^0 \times \mathbb{R}_+$ tais que*

$$r := \frac{1}{\tilde{\lambda}} \nabla(dw)_z(z_-) \in \partial_\varepsilon(\hat{\Phi}_{\tilde{z}})(\tilde{z}), \quad (5.2)$$

$$(dw)_z(\tilde{z}) + \tilde{\lambda}\varepsilon \leq \sigma(dw)_{z_-}(\tilde{z}), \quad (5.3)$$

com $\hat{\Phi}_{\tilde{z}}$ dado por (1.14).

Além da Seção 5.1, este capítulo contém a Seção 5.2 na qual descrevemos uma versão acelerada do Algoritmo 10 para resolver $\text{SP}(\hat{\Phi}; Z)$ e fornecemos o correspondente resultado de complexidade. Também discutimos modos ótimos de escolher o tamanho do passo de modo a minimizar a complexidade total da versão.

5.1 UM ESQUEMA ACELERADO PARA RESOLVER (P2)

Nesta subseção apresentamos um esquema para determinar uma solução do problema (P2) baseado no algoritmo de gradiente acelerado da Seção 1.4 aplicado a um certo problema de min-max relacionado.

Tendo em vista este objetivo, considere o problema de min-max regularizado

$$\min_{x \in X} \max_{y \in Y} \hat{\Phi}(x, y) + \frac{1}{\lambda}(dw^X)_{x_-}(x) - \frac{1}{\lambda}(dw^Y)_{y_-}(y). \quad (5.4)$$

Podemos verificar que uma solução exata de (5.4) determina uma solução de (P2) com

$\sigma = 0$ na qual $\tilde{\lambda} = \lambda$. De fato, fazendo $\sigma = 0$ em (5.3), ficamos com $\varepsilon = 0$ e $z = \tilde{z}$. Substituindo em (5.2) ficamos com $\frac{1}{\lambda} \nabla dw_z(z_-) \in \partial \hat{\Phi}_z(z)$, que é a condição de otimalidade de (5.4). Sendo assim, é razoável se esperar que soluções aproximadas de (5.4) resultem em soluções aproximadas de (P2).

Fazendo

$$f_\lambda(x) := \max_{y \in Y} \left\{ \Phi(x, y) - \frac{1}{\lambda} (dw^Y)_{y_-}(y) \right\}, \quad \forall x \in X, \quad (5.5)$$

$$g_\lambda(x) := \frac{1}{\lambda} (dw^X)_{x_-}(x) + \phi(x), \quad \forall x \in \mathcal{X}, \quad (5.6)$$

segue de (2.2), (5.5) e (5.6) que (5.4) é equivalente a (1.24) com $(f, g) = (f_\lambda, g_\lambda)$. Além disso, as condições A.1 and A.2 são satisfeitas com $\mu = \eta_X / \lambda$ por (5.5) e pelo fato de que dw^X é uma (η_X, X) -distância de Bregman sobre W_X . O seguinte resultado estabelece a validade de A.3.

Proposição 5.1 *A constante $L = L_\lambda$ e a função $\nabla f = \nabla f_\lambda : X \rightarrow X$ definidas como*

$$L_\lambda := 2 \left(L_{xx} + \frac{\lambda}{\eta_Y} L_{xy}^2 \right), \quad \nabla f_\lambda(x) := \nabla_x \Phi(x, y_\lambda(x)), \quad \forall x \in X, \quad (5.7)$$

respectivamente, onde $y_\lambda(x)$ é definido como

$$y_\lambda(x) := \operatorname{argmax}_{y \in Y} \left\{ \Phi(x, y) - \frac{1}{\lambda} (dw^Y)_{y_-}(y) \right\}, \quad \forall x \in X, \quad (5.8)$$

satisfazem a condição A.3 com $f = f_\lambda$.

Prova. A prova é análoga à da Proposição 2.1.

■

A seguir apresentaremos o algoritmo para resolver (P2) supondo que o dado de entrada z_- está em $W^0 \cap Z$. Este algoritmo consiste em aplicar o algoritmo de gradiente

acelerado da Seção 1.4 ao problema (1.24) com $(f, g) = (f_\lambda, g_\lambda)$, onde f_λ e g_λ são dadas como em (5.5) e (5.6), respectivamente.

Algoritmo 11: Algoritmo acelerado para resolver (P2).

Passo 0 Dados $\sigma \in (0, 1]$, $\lambda > 0$, $\tau \in (0, 1)$ e $z_- = (x_-, y_-) \in W^0 \cap Z$, faça $A_0 = 0$, $k = 1$,

$$\tilde{\Lambda}_0 \equiv 0, \tilde{y}_0 = 0, L_\lambda \text{ como em (5.7), e } x_0 = \tilde{x}_0 := x_-;$$

Passo 1 compute

$$A_k := A_{k-1} + \frac{(1 + \mu A_{k-1}) + \sqrt{(1 + \mu A_{k-1})^2 + 4L_\lambda(1 + \mu A_{k-1})A_{k-1}}}{2L_\lambda},$$

com $\mu = \eta_X/\lambda$, calcule

$$\check{x}_k := \frac{A_{k-1}}{A_k} \tilde{x}_{k-1} + \frac{A_k - A_{k-1}}{A_k} x_{k-1},$$

compute $y_\lambda(\check{x}_k)$ de acordo com (5.8), e a função afim $\tilde{\Lambda}_k$ como

$$\tilde{\Lambda}_k := \frac{A_{k-1}}{A_k} \tilde{\Lambda}_{k-1} + \frac{A_k - A_{k-1}}{A_k} [\Phi(\check{x}_k, y_\lambda(\check{x}_k)) + \langle \nabla \Phi(\check{x}_k, y_\lambda(\check{x}_k)), \cdot - \check{x}_k \rangle_{\mathcal{X}}]; \quad (5.9)$$

Passo 2 faça

$$\lambda_k = \left(\frac{1}{\lambda} + \frac{1}{\eta_X A_k} \right)^{-1}, \quad (5.10)$$

e compute os iterados x_k , \tilde{y}_k e \tilde{x}_k dados por

$$x_k = \operatorname{argmin} \left\{ \tilde{\Lambda}_k(x) + \phi(x) + \frac{1}{\lambda_k} (dw^X)_{x_-}(x) \right\}, \quad (5.11)$$

$$\tilde{y}_k = \frac{A_{k-1}}{A_k} \tilde{y}_{k-1} + \frac{A_k - A_{k-1}}{A_k} y_\lambda(\check{x}_k), \quad (5.12)$$

$$\tilde{x}_k := \frac{A_{k-1}}{A_k} \tilde{x}_{k-1} + \frac{A_k - A_{k-1}}{A_k} x_k;$$

Passo 3 se $\lambda_k \geq \max\{1 - \sigma, \tau\}\lambda$, então compute $y_k := y_{\lambda_k}(\tilde{x}_k)$ de acordo com (5.8), faça

$$\tilde{\lambda} = \lambda_k, \tilde{z} = \tilde{z}_k := (\tilde{x}_k, \tilde{y}_k), z = z_k := (x_k, y_k) \text{ e}$$

$$\varepsilon = \varepsilon_k := \hat{\Phi}(\tilde{x}_k, y_k) - \tilde{\Lambda}_k(x_k) - \phi(x_k) - \frac{1}{\lambda_k} \langle \nabla(dw)_z(z_-), \tilde{z} - z \rangle,$$

retorne $\tilde{\lambda}$ e a tripla $(\tilde{z}, z, \varepsilon)$, e termine o algoritmo; caso contrário, faça $k \leftarrow k + 1$ e retorne ao Passo 1.

fim

A seguir fazemos algumas considerações acerca do Algoritmo 11. Primeiramente, pelo critério de parada e pela relação (5.10), o Algoritmo 11 retorna $\tilde{\lambda} \in [\tau\lambda, \lambda]$. Em segundo lugar, pela Proposição 4.1 e pelas relações (2.1), (5.8) e (5.11), a saída z está em $W^0 \cap Z$. Em terceiro lugar, os passos 1 e 2 do Algoritmo 11 são especializações dos passos 1 e 2 do Algoritmo 1 para a instância de (1.24) na qual (f, g) é dado por (f_λ, g_λ) com f_λ e g_λ como em (5.5) e (5.6), respectivamente. A única diferença é a computação extra de \tilde{y}_k em (5.12), que é usado para computar a componente \tilde{z} da saída. Em quarto lugar, mesmo que a função afim $\tilde{\Lambda}_k$ dada em (5.9) e a função afim Λ_k dada por (1.26) com $f = f_\lambda$ não sejam a mesma, ambas têm o mesmo gradiente por causa de (5.7), e portanto os subproblemas (5.11) e (1.27) são equivalentes. Em quinto lugar, cada iteração do Algoritmo 11 antes da última requer resolver dois subproblemas, a saber, (5.11) e um da forma (5.8), enquanto a última iteração requer a resolução de um subproblema adicional da forma (5.8) no Passo 3. Em sexto lugar, quando o critério de parada no Passo 3 é satisfeito, este passo extra computa a saída $\tilde{\lambda}$ e a tripla $(\tilde{z}, z, \varepsilon)$ resolvendo (P2) (veja a Proposição 5.2 a seguir). Por último, outro modo possível de terminar o Algoritmo 11 é computar a tripla $(\tilde{z}, z, \varepsilon) = (\tilde{z}_k, z_k, \varepsilon_k)$ descrita no Passo 3 a cada iteração e checar se $\tilde{\lambda} = \lambda_k$ e se esta tripla satisfaz o critério de erro (5.3). (A tripla sempre satisfaz (5.2) por causa da Proposição 5.2 (a) enunciada a seguir). O lado ruim deste critério de parada é que ele requer resolver um subproblema adicional da forma (5.8) a cada iteração. Nossos testes numéricos apresentados no Capítulo 6 são baseados no critério de parada descrito

no Algoritmo 11.

O seguinte resultado estabelece a convergência e complexidade do Algoritmo 11.

Proposição 5.2 *Para todo $k \geq 1$ valem as seguintes propriedades:*

- (a) *o escalar $\tilde{\lambda} = \lambda_k$ e a tripla $(\tilde{z}, z, \varepsilon) = (\tilde{z}_k, z_k, \varepsilon_k)$ satisfazem a inclusão (5.2);*
- (b) *se $\lambda_k \geq (1 - \sigma)\lambda$, então $\tilde{\lambda} = \lambda_k$ e a tripla $(\tilde{z}, z, \varepsilon) = (\tilde{z}_k, z_k, \varepsilon_k)$ satisfazem a condição (5.3).*

Além disso, o Algoritmo 11 resolve o problema (P2) em no máximo

$$\mathcal{O} \left(\left\lceil \sqrt{\frac{\lambda \left(L_{xx} + \frac{\lambda}{\eta_Y} L_{xy}^2 \right)}{\eta_X}} \right\rceil \right) \quad (5.13)$$

iterações.

Antes de provar esta proposição precisamos do seguinte resultado auxiliar:

Lema 5.1 *Para todo $k \geq 1$, a função afim $\tilde{\Lambda}_k$ dada por (5.9) satisfaz:*

- (a) *$\tilde{\Lambda}_k(x) \leq \Phi(x, \tilde{y}_k)$ para todo $x \in X$;*
- (b) *fazendo $\tilde{z}_k := (\tilde{x}_k, \tilde{y}_k)$, a função $\Gamma_{\tilde{z}_k} : \mathcal{Z} \rightarrow [-\infty, \infty]$ definida como*

$$\Gamma_{\tilde{z}_k}(z) := \begin{cases} \tilde{\Lambda}_k(x) + \phi(x) - \hat{\Phi}(\tilde{x}_k, y), & \forall z = (x, y) \in Z; \\ +\infty, & \text{caso contrário,} \end{cases} \quad (5.14)$$

minoriza a função $\hat{\Phi}_{\tilde{z}_k}$ definida em (1.14);

- (c) *se $\lambda_k \geq (1 - \sigma)\lambda$ então*

$$\frac{(1 - \sigma)}{\lambda_k} (dw)_{(z_-)}(\tilde{z}_k) \leq \inf_{u \in \mathcal{Z}} \left\{ \Gamma_{\tilde{z}_k}(u) + \frac{1}{\lambda_k} (dw)_{(z_-)}(u) \right\}. \quad (5.15)$$

Prova. (a) Usando as definições de $\tilde{\Lambda}_k$ e \tilde{y}_k dadas em (5.9) e (5.11) e também o fato de que $\Phi(\cdot, y) - \Phi(x, \cdot)$ é convexa para todo $(x, y) \in Z$, vemos que para todo $x \in X$

$$\begin{aligned}\tilde{\Lambda}_k(x) &= \sum_{i=1}^k \frac{A_i - A_{i-1}}{A_k} [\Phi(\check{x}_i, y_\lambda(\check{x}_i)) + \langle \nabla \Phi(\check{x}_i, y_\lambda(\check{x}_i)), x - \check{x}_i \rangle_{\mathcal{X}}] \\ &\leq \sum_{i=1}^k \frac{A_i - A_{i-1}}{A_k} [\Phi(x, y_\lambda(\check{x}_i))] \leq \Phi\left(x, \sum_{i=1}^k \frac{A_i - A_{i-1}}{A_k} y_\lambda(\check{x}_i)\right) = \Phi(x, \tilde{y}_k),\end{aligned}$$

o que prova (a).

(b) por (1.14), (2.2) e (5.14), vemos que $\Gamma_{\tilde{z}_k}$ minoriza $\hat{\Phi}_{\tilde{z}_k}$ se e somente se $\tilde{\Lambda}_k \leq \Phi(\cdot, \tilde{y}_k)$, e portanto segue a prova desse item.

(c) Suponha que $\lambda_k \geq (1 - \sigma)\lambda$. Observe que por (5.5), (5.7), (5.8), (5.9), (5.12) e pela convexidade de $(dw^Y)_{y_-}(\cdot)$, temos que

$$\begin{aligned}\tilde{\Lambda}_k(x) - \frac{1}{\lambda}(dw^Y)_{y_-}(\tilde{y}_k) &\geq \sum_{i=1}^k \frac{A_i - A_{i-1}}{A_i} \left[\Phi(\check{x}_i, y_\lambda(\check{x}_i)) + \langle \nabla \Phi(\check{x}_i, y_\lambda(\check{x}_i)), x - \check{x}_i \rangle_{\mathcal{X}} - \frac{1}{\lambda}(dw^Y)_{y_-}(y_\lambda(\check{x}_i)) \right] \\ &= \sum_{i=1}^k \frac{A_i - A_{i-1}}{A_i} [f_\lambda(\check{x}_i) + \langle \nabla f_\lambda(\check{x}_i), x - \check{x}_i \rangle_{\mathcal{X}}].\end{aligned}\tag{5.16}$$

Agora, fazendo $x_0 = x_-$, $g = g_\lambda$ e $h = (1/\eta_X)dw^X_{x_-}$ onde g_λ é dada por (5.6), e usando o fato de que dw^X é uma (η_X, X) -distância de Bregman sobre W , verificamos que x_0 , g e h satisfazem as condições A.4-A.6. Uma vez que o conjunto X e o par $(f, g) = (f_\lambda, g_\lambda)$ satisfazem A.1-A.3 onde f_λ é dada por (5.5), e o Algoritmo 11 corresponde ao Algoritmo 1 aplicado a (1.24) com $(f, g) = (f_\lambda, g_\lambda)$ com h dada como acima. Segue de (1.32), (5.5), (5.6) (5.10) e (5.16) que

$$\begin{aligned}\tilde{\Lambda}_k(x) + \phi(x) - \frac{1}{\lambda}(dw^Y)_{y_-}(\tilde{y}_k) + \frac{1}{\lambda_k}(dw^X)_{x_-}(x) &= \tilde{\Lambda}_k(x) + \phi(x) - \frac{1}{\lambda}(dw^Y)_{y_-}(\tilde{y}_k) + \left(\frac{1}{\lambda} + \frac{1}{\eta_X A_k}\right)(dw^X)_{x_-}(x) \\ &\geq \sum_{i=1}^k \frac{A_i - A_{i-1}}{A_i} [f_\lambda(\check{x}_i) + \langle \nabla f_\lambda(\check{x}_i), x - \check{x}_i \rangle_{\mathcal{X}}] + g_\lambda(x) + \frac{1}{A_k}h(x) \\ &\geq (f_\lambda + g_\lambda)(\tilde{x}_k) \geq \hat{\Phi}(\tilde{x}_k, y) - \frac{1}{\lambda}(dw^Y)_{y_-}(y) + \frac{1}{\lambda}(dw^X)_{x_-}(\tilde{x}_k), \quad \forall (x, y) \in Z.\end{aligned}$$

Por fim, rearranjando a última desigualdade e usando as definições de dw e $\Gamma_{\tilde{z}_k}$ em (5.1) e (5.14), respectivamente, e também o fato de que $(1 - \sigma)\lambda \leq \lambda_k \leq \lambda$, onde a segunda desigualdade deve-se a (5.10), temos que vale (5.15).

■

Agora estamos prontos para provar a Proposição 5.2.

Prova. [**P**rova da Proposição 5.2] Considere $k \geq 1$. A prova dos dois primeiros itens é baseada na Proposição 4.3 considerando-se $\lambda = \lambda_k$, $\tilde{z} = \tilde{z}_k$, $\varepsilon = \varepsilon_k$ e o operador $T = T_{\hat{\Phi}}$ dado por (1.15), que satisfaz a condição D.1 com $f_{\tilde{z}} = \hat{\Phi}_{\tilde{z}}$ dada por (1.14) para todo $\tilde{z} \in Z$ (ver a observação abaixo da condição D.1).

(a) o Lema 5.1 (b) implica que $\Gamma_{\tilde{z}_k}$ definida como em (5.14) minoriza $\hat{\Phi}_{\tilde{z}_k}$. Sendo assim, usando o fato de que $f_{\tilde{z}}(\tilde{z}) = \hat{\Phi}_{\tilde{z}}(\tilde{z}) = 0$ para todo $\tilde{z} \in Z$ pela definição (1.14), vemos que as grandezas z e ε computadas de acordo com (4.15) e (4.16), respectivamente, com $\tilde{z} = \tilde{z}_k$ é equivalente ao modo que z_k e ε_k são computados no Algoritmo 11. Portanto, (a) segue da Proposição 4.3 (a).

(b) Este item segue pelos mesmos argumentos acima, Lema 5.1 (c) e Proposição 4.3 (b).

Agora estabelecemos a última conclusão da proposição. Quando o Algoritmo 11 termina (ver o Passo 3), segue de (5.10) e do fato de que a saída $\tilde{\lambda} = \lambda_k$ satisfaz $\tilde{\lambda} \in [\tau\lambda, \lambda]$. Portanto, pelos itens (a) e (b), concluímos que o Algoritmo 11 retorna $\tilde{\lambda}$ e $(\tilde{z}, z, \varepsilon)$ resolvendo (P2). Falta verificar a estimativa (5.13). De fato, vamos estimar o número de iterações para que o Algoritmo 11 termine. Ou seja, queremos que $\lambda_k \geq \max\{1 - \sigma, \tau\}\lambda$. Por (5.10) isto se dá se e somente se

$$\frac{1}{\lambda} + \frac{1}{\eta_X A_k} \leq \frac{\max\{1 - \sigma, \tau\}}{\lambda} \Leftrightarrow \frac{1}{\eta_X A_k} \leq \frac{1}{\lambda} [\max\{1 - \sigma, \tau\} - 1] := \frac{c_{\sigma\tau}}{\lambda}.$$

Lembrando a estimativa $A_k \geq k^2/4L_\lambda$ dada em (1.30), temos que para atingir o critério de parada, basta que k seja tal que

$$\frac{4L_\lambda}{\eta_X k^2} \leq \frac{c_{\sigma\tau}}{\lambda},$$

o que com (5.7) resulta na estimativa (5.13). Verifica-se, portanto, que o número de iterações até o critério de parada do Algoritmo 11 ser satisfeito é limitado por (5.13) quando τ e σ são vistos como constantes tais que $\max\{1 - \sigma, \tau\}$ é suficientemente diferente de 1 e de 0.

■

5.2 UMA VERSÃO ACELERADA NE-HPE PARA RESOLVER $\text{SP}(\hat{\Phi}; Z)$

Nesta subseção descrevemos uma versão acelerada NE-HPE para resolver o problema de ponto-de-sela $\text{SP}(\hat{\Phi}; Z)$ e estabelecemos seu correspondente resultado de complexidade. Também discutimos a maneira ótima de escolher o tamanho do passo de modo a minimizar a complexidade total da versão.

Começamos enunciando uma versão acelerada do Algoritmo 10 para resolver $\text{SP}(\hat{\Phi}; Z)$ que computa o tamanho do passo λ_j e a tripla $(\tilde{z}_j, z_j, \varepsilon_j)$ no Passo 1 fazendo uso do Algoritmo 11.

Algoritmo 12: Algoritmo acelerado NE-HPE para o problema de ponto-de-sela

Passo 0 tome $z_0 \in W^0$, $\lambda > 0$, $\sigma \in (0, 1]$ e $\tau \in (0, 1)$ e faça $j = 1$;

Passo 1 invoque o Algoritmo 11 com dados de entrada σ , λ , τ e $z_- = z_{j-1}$ para obter um

tamanho do passo λ_j e uma tripla $(\tilde{z}_j, z_j, \varepsilon_j)$ satisfazendo (4.11) e (4.5);

Passo 2 faça $j \leftarrow j + 1$, e retorne ao Passo 1.

fm

Em vista da Proposição 5.2, o Algoritmo 12 satisfaz as condições de erro (4.11) e (4.5) do Passo 1 do Algoritmo 10. Portanto, o Algoritmo 12 é um caso especial do Algoritmo 10. Segue que o limitante de taxa de convergência ergódica (externo) para o Algoritmo 12 é como descrito no Teorema 4.1.

Teorema 5.1 *Considere $R = R(z_0; Z)$ dado por (4.8). Considere as sequências $\{\tilde{z}_j\}$, $\{z_j\}$ e $\{\varepsilon_j\}$ geradas pelo Algoritmo 12 e as respectivas sequências ergódicas $\{\tilde{z}_j^a\}$, $\{z_j^a\}$ e $\{\varepsilon_j^a\}$ como no Lema 4.4. Então, valem:*

(a) *para todo escalar positivo $\bar{\varepsilon}$, existe um índice*

$$j_0 = \mathcal{O}\left(\left\lceil \frac{R}{\lambda \bar{\varepsilon}} \right\rceil\right)$$

tal que para todo $j \geq j_0$, $(\tilde{z}_j^a, \varepsilon_j^a)$ é um $\bar{\varepsilon}$ -ponto-de-sela de $\hat{\Phi}$ com respeito a Z ;

(b) *cada iteração do Algoritmo 12 performa no máximo*

$$\mathcal{O}\left(\left\lceil \sqrt{\frac{\lambda \left(L_{xx} + \frac{\lambda}{\eta_Y} L_{xy}^2\right)}{\eta_X}} \right\rceil\right)$$

iterações internas, isto é, do Algoritmo 11.

Como consequência, o Algoritmo 12 determina um $\bar{\varepsilon}$ -ponto-de-sela de $\hat{\Phi}$ com respeito a Z performando no máximo

$$\mathcal{O}\left(\left\lceil \sqrt{\frac{\lambda \left(L_{xx} + \frac{\lambda}{\eta_Y} L_{xy}^2\right)}{\eta_X}} \right\rceil \left\lceil \frac{R}{\lambda \bar{\varepsilon}} \right\rceil\right) \quad (5.17)$$

iterações internas.

Prova. Uma vez que o Algoritmo 12 é uma versão especial do Algoritmo 10, (a) segue do Teorema 4.1 (a) e do fato de que $\lambda_j \geq \tau\lambda$ para todo $j \geq 1$. O item (b) segue da Proposição 5.2. A última afirmação do teorema segue imediatamente de (a) e (b). ■

Terminamos esta subseção fazendo uma consideração acerca do limite de complexidade (5.17) em comparação àquele obtido na relação (4.4) de (NESTEROV, 2005). Quando substituimos $\lambda = R/\bar{\varepsilon}$ em (5.17), verificamos que o limite (5.17) se reduz a

$$\mathcal{O} \left(1 + \frac{RL_{xy}}{\bar{\varepsilon}\sqrt{\eta_X\eta_Y}} + \sqrt{\frac{RL_{xx}}{\bar{\varepsilon}\eta_X}} \right). \quad (5.18)$$

A seguir mostramos que, para distâncias de Bregman com respeito a X e Y convenientemente escalonadas, este limite se reduz à

$$\mathcal{O} \left(1 + \frac{\sqrt{R_X R_Y} L_{xy}}{\bar{\varepsilon}\sqrt{\eta_X\eta_Y}} + \sqrt{\frac{R_X L_{xx}}{\bar{\varepsilon}\eta_X}} \right), \quad (5.19)$$

onde

$$R_X := \max\{(dw^X)_{x_0}(x) : x \in X\}, \quad R_Y := \max\{(dw^Y)_{y_0}(y) : y \in Y\}.$$

O último limite generaliza aquele na relação (4.4) de (NESTEROV, 2005), o qual é válido apenas para um caso especial de $\text{SP}(\hat{\Phi}; Z)$ com uma estrutura bilinear. Este limitante é ótimo no seguinte sentido:

- a) Pode ser mostrado que existe uma função G com gradiente Lipschitz de constante L_G tal que para qualquer método de primeira ordem a taxa de convergência para minimizar G em um conjunto X é de no máximo $\mathcal{O}(L_G/k^2)$. (NESTEROV, 2004)
- b) Pode ser mostrado que existem $Y' \subset \mathbb{R}^n$ convexo e compacto, $b \in Y'$ e um operador linear limitado K tais que, para qualquer método de primeira ordem, a taxa de convergência para resolver $\min_{x \in X} \max_{y \in Y'} \langle Kx, y \rangle - J(y) := \min_{x \in X} \max_{y \in Y'} \langle Kx - b, y \rangle$ é de no máximo $\mathcal{O}(L_K/k)$. (NEMIROVSKI, 2004; NEMIROVSKY, 1992)

A fim de obter o limite (5.19), considere as distâncias de Bregman definidas como

$$dw^{X,\theta} := \theta dw^X, \quad dw^{Y,\theta} := \theta^{-1} dw^Y,$$

onde $\theta > 0$ é um parâmetro fixado. Claramente, $dw^{X,\theta}$ (respectivamente, $dw^{Y,\theta}$) é uma $(\theta\eta_X, X)$ -distância de Bregman (respectivamente, $(\theta^{-1}\eta_Y, Y)$ -distância de Bregman) sobre W_X (respectivamente, sobre W_Y). Neste caso, R , dado na equação (4.8), torna-se

$$R = R_\theta := \theta R_X + \theta^{-1} R_Y.$$

Sendo assim, escolhendo $\theta = (R_Y/R_X)^{1/2}$, as grandezas R , η_X e η_Y neste caso se reduzem a

$$R = 2\sqrt{R_X R_Y}, \quad \eta_X = \sqrt{\frac{R_Y}{R_X}} \eta_X, \quad \eta_Y = \sqrt{\frac{R_X}{R_Y}} \eta_Y,$$

e portanto (5.18) se reduz a (5.19).

Sob certas condições, podemos escolher λ ainda menor que $R/\bar{\varepsilon}$ de modo a manter a mesma complexidade. Exporemos tais condições a seguir. Multiplicando a estimativa (5.17) por $\lambda(1/\lambda)$, verificamos que ela é equivalente a

$$\mathcal{O} \left(\max \left\{ \frac{1}{\lambda}, \sqrt{\frac{L_{xx}}{\lambda \eta_X}}, \frac{L_{xy}}{\sqrt{\eta_X \eta_Y}} \right\} \max \left\{ \lambda, \frac{R}{\bar{\varepsilon}} \right\} \right).$$

Podemos ver então que a afirmação a seguir é válida dependendo se a desigualdade

$$\lambda_1 := \max \left\{ \frac{\sqrt{\eta_X \eta_Y}}{L_{xy}}, \frac{L_{xx}}{L_{xy}^2 \eta_Y} \right\} \leq \frac{R}{\bar{\varepsilon}} =: \lambda_2 \quad (5.20)$$

vale (caso 1) ou não (caso 2):

1) se (5.20) vale, então escolhendo $\lambda \in [\lambda_1, \lambda_2]$ a complexidade dada em (5.17) se reduz

a

$$\mathcal{O} \left(\frac{L_{xy}}{\sqrt{\eta_X \eta_Y}} \frac{R}{\bar{\varepsilon}} \right);$$

2) caso contrário, se $\lambda_1 > \lambda_2$, então tomando $\lambda = \lambda_2$ a complexidade dada em (5.17) se

reduz a

$$\mathcal{O} \left(1 + \sqrt{\frac{L_{xx}}{\eta_X}} \sqrt{\frac{R}{\bar{\varepsilon}}} \right).$$

Fazendo as mesmas modificações nas distâncias de Bregman, obtemos assim a mesma complexidade em (5.19) com esta escolha de λ .

Idealmente, deveríamos escolher λ de acordo com a discussão acima para minimizar o número total de iterações. Contudo, no caso em que R não é conhecido a priori, não podemos computar λ_2 , e com isso não podemos escolher $\lambda = \lambda_2$ como proposto no caso 2. Por outro lado, sempre podemos escolher $\lambda = \lambda_1$, que é facilmente computável.

Esta escolha é ótima quando vale o caso 1 e, mesmo que não o seja quando vale o caso 2, pode ser uma escolha boa na prática, visto que o caso 2 é difícil de acontecer.

6 EXPERIMENTOS NUMÉRICOS

Neste capítulo, apresentamos resultados computacionais ilustrando a performance numérica do Algoritmo 12 em uma coleção de problemas de ponto-de-sela. Todos os resultados computacionais foram programados em MATLAB R2014a em uma máquina Windows 64 bit com processador Intel 2.16 GHz de 4 GB de memória.

Comparamos o Algoritmo 12 (referido como ACC-HPE) com o algoritmo de suavização de Nesterov ([NESTEROV, 2005](#)) (referido como NEST). Implementamos ambos os algoritmos baseados tanto na distância Euclidiana quanto na distância de Bregman induzida pela divergência de Kulback-Leibler, a saber, $dw_{z^2}(z^1) = \sum_i z_i^1 \log(z_i^1/z_i^2) + z_i^1 - z_i^2$. Nossos resultados computacionais consideram então quatro variantes, que são, E-ACC-HPE, L-ACC-HPE, E-NEST e L-NEST, onde aquelas começando com E (respectivamente, L) são as baseadas na distância Euclidiana (respectivamente, distância de Kulback-Leibler). Para melhorar a performance das variantes L , usamos o esquema adaptativo para escolher o parâmetro L dado em ([TSENG, 2008](#)), *i.e.*, o valor inicial de L começa como uma fração do valor teórico da constante de Lipschitz e é aumentado de 2 sempre que falha em satisfazer um certo critério de convergência (veja as equações (23) e (45) de ([TSENG, 2008](#))). Usamos a fração $1/2^9$ em nossos experimentos. Este esquema não foi usado para as variantes E, dado que observamos que ele não melhora suas performances. O valor retornado de L na última iteração dividido pelo valor teórico da constante de Lipschitz variou entre $1/64$ e 1 em nossos experimentos. Mais especificamente, esta razão foi de $1/64$ para uma instância, $1/32$ para três instâncias, $1/8$ para uma instância, $1/4$ para quatro instâncias, $1/2$ para treze instâncias e 1 para as demais instâncias.

Nas três seguintes subseções reportamos resultados computacionais nas seguintes

classes de problemas: (a) problema bilinear; (b) problema de ponto-de-sela vetor-matriz; e (c) problema quadrático. Reportamos os resultados em tabelas e em perfis de desempenho (ver (DOLAN; MORÉ, 2002)). Lembramos a definição de perfil de desempenho: dado um problema, um método A é dito ser no máximo x vezes mais devagar que o método B se o tempo que o método A gastou para resolvê-lo é no máximo x vezes o tempo gasto pelo método B . O ponto (x, y) está na curva do perfil de desempenho de um método se este resolver exatamente $100y\%$ de todos os problemas testados em até, na pior das hipóteses, x vezes mais devagar que qualquer outro método em comparação.

Para todas as classes de problemas, o critério de parada utilizado para terminar todos os métodos na k -ésima iteração é

$$\max_{y \in Y} \hat{\Phi}(\tilde{x}_k, y) - \min_{x \in X} \hat{\Phi}(x, \tilde{y}_k) \leq \bar{\varepsilon}.$$

O uso deste critério para a segunda e terceira classes de problemas não é a melhor estratégia do ponto de vista computacional, já que a minimização da função dual envolve resolver um subproblema quadrático no simplex unitário. Note que nosso algoritmo nos permite computar a cada iteração um par $((\tilde{x}_k, \tilde{y}_k), \varepsilon_k)$ tal que a desigualdade acima é satisfeita com $\bar{\varepsilon} = \varepsilon_k$, e assim o critério de parada acima será satisfeito sempre que $\varepsilon_k \leq \bar{\varepsilon}$. No entanto, uma vez que a descrição do algoritmo de suavização de Nesterov gera $(\tilde{x}_k, \tilde{y}_k)$ mas não ε_k , optamos pelo critério mencionado acima mas adotamos a convenção de excluir o tempo gasto pelos programas de avaliar as funções duais para os tempos reportados.

Selecionamos os parâmetros para nossos algoritmos como a seguir. Para o Algoritmo 12 escolhemos

$$\lambda = \min \left\{ R/\varepsilon, \max \left\{ \frac{\sqrt{\eta_X \eta_Y}}{L_{xy}}, \frac{L_{xx}}{L_{xy}^2 \eta_Y} \right\} \right\},$$

conforme a discussão no fim da Seção 5.2, $\sigma = 0.9$ e $\tau = 1 - \sigma$. Escolhemos o ponto inicial sempre como $(1/n, 1/n, \dots, 1/n)$ onde n é a dimensão do espaço considerado. Em

nossa implementação do algoritmo de suavização de Nesterov escolhemos o parâmetro de suavização $\mu = \varepsilon/2D_Y$ (ver a equação (4.8) de (NESTEROV, 2005)).

A seguir denote por \mathbb{R}^n o espaço Euclidiano n -dimensional e \mathcal{S}^n o espaço linear das matrizes simétricas $n \times n$ reais. O simplex unitário em \mathbb{R}^n é definido como

$$\Delta_n := \left\{ x \in \mathbb{R}^n : \sum_{i=1}^n x_i = 1, x \geq 0 \right\}.$$

6.1 PROBLEMA BILINEAR

Nesta seção exibimos os resultados comparando a performance das quatro variantes em instâncias do problemas bilinear

$$\min_{x \in \Delta_n} \max_{y \in \Delta_m} \langle x, Ay \rangle,$$

onde A é uma matriz $n \times m$ real. As matrizes foram geradas de modo que seus elementos são não nulos com probabilidade p e randomicamente gerados no intervalo $[-1, 1]$. Testamos os algoritmos para um conjunto de problemas de tamanhos diferentes de matrizes e diferentes valores de p . A tolerância usada aqui foi $\bar{\varepsilon} = 10^{-3}$.

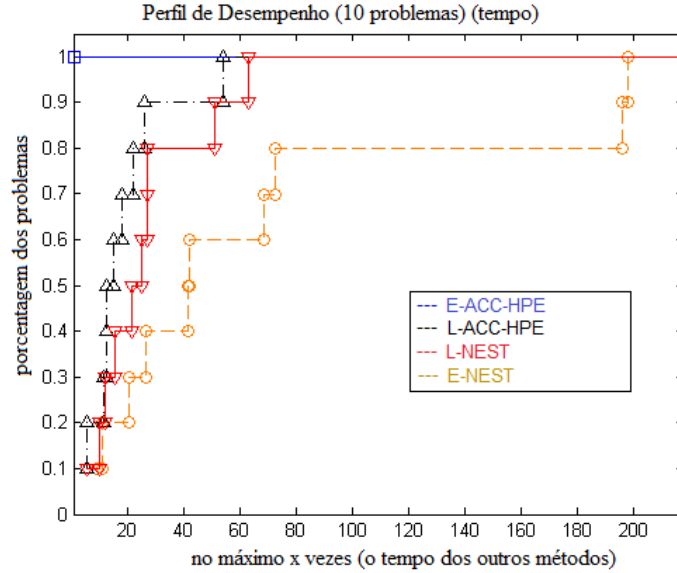
Na tabela 1 vemos os resultados para as quatro variantes aplicados a instâncias deste problema com diversos tamanhos de matrizes e diversos valores de p .

TABELA 1 – Resultados dos testes para o problema bilinear

Tamanho			E-ACC-HPE		E-NEST		L-ACC-HPE		L-NEST	
n	m	p	tempo	iter.	tempo	iter.	tempo	iter.	tempo	iter.
1000	100	0.01	0.2721	196	1.8915	1806	6.0765	4364	6.0991	4979
1000	100	0.1	0.6587	480	14.5410	12738	8.88550	5734	7.7157	5808
1000	1000	0.01	0.4562	224	3.9617	2560	7.1988	1245	30.0177	2287
1000	1000	0.1	0.7927	213	47.3204	14602	12.2948	2723	17.1655	4603
1000	10000	0.01	0.7082	100	28.0016	4213	12.9047	1799	83.6426	5098
1000	10000	0.1	3.5575	196	698.2147	38410	67.6110	2541	259.4104	5306
10000	100	0.01	1.8140	461	33.8041	9606	36.3373	7010	33.5731	7245
10000	100	0.1	11.5663	1381	745.7268	100384	48.0043	7671	47.5234	7740
10000	1000	0.01	0.7976	121	34.0278	5038	62.5932	6572	67.1911	7608
10000	1000	0.1	8.1079	287	1566.3	56744	165.6846	7449	161.6192	7816

Na figura 3 vemos o perfil de desempenho para o mesmo conjunto de instâncias. Ela mostra que as variantes do ACC-HPE são superiores às variantes NEST neste conjunto de instâncias do problema bilinear.

FIGURA 2 – Perfil de desempenho para o problema bilinear



6.2 PROBLEMA DE PONTO-DE-SELA VETOR-MATRIZ

Nesta seção vemos a comparação das quatro variantes de instâncias do problema de ponto-de-sela vetor-matriz. Dados $c \in \mathbb{R}^n$, uma matriz B $n \times n$ real e um operador linear $\mathcal{A} : \mathbb{R}^n \rightarrow \mathcal{S}^m$, o problema de ponto-de-sela vetor-matriz consiste em

$$\min_{x \in \Delta_n} \frac{1}{2} \|Bx + c\|^2 + \theta_{\max}(\mathcal{A}(x)),$$

onde $\theta_{\max}(\mathcal{A}(x))$ denota o maior autovalor de $\mathcal{A}(x)$. Tal problema é equivalente ao problema de ponto-de-sela

$$\min_{x \in \Delta_n} \max_{Y \in \Omega} \frac{1}{2} \|Bx + c\|^2 + \langle \mathcal{A}(x), Y \rangle,$$

onde $\Omega := \{Y \in \mathcal{S}^m : \text{tr}(Y) = 1, Y \text{ é positiva definida}\}$. Testamos as quatro variantes em um conjunto de problemas onde as matrizes B e $\mathcal{A}_i := \mathcal{A}(e_i)$, $i = 1, \dots, n$, foram geradas de modo que seus elementos são não nulos com probabilidade 0.1 e randomicamente

gerados no intervalo $[-1, 1]$. (Aqui, e_i denota a i -ésima componente do vetor n -dimensional unitário). A tolerância utilizada foi $\bar{\varepsilon} = 10^{-2}$.

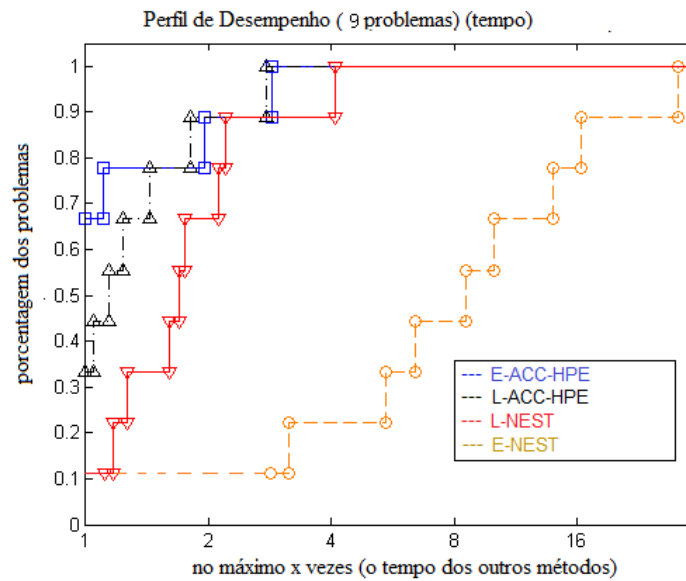
Na tabela 2 vemos os resultados para as quatro variantes aplicado a diversas instâncias deste problema para diversos tamanhos de matrizes.

TABELA 2 – Resultados dos testes para o problema de ponto-de-sela vetor-matriz

Tamanho		E-ACC-HPE		E-NEST		L-ACC-HPE		L-NEST	
n	m	tempo	iter.	tempo	iter.	tempo	iter.	tempo	iter.
50	50	1.5987	266	3.9912	750	1.2079	240	1.7442	300
50	100	10.4205	456	44.2678	2335	3.7254	235	7.9121	410
50	200	56.4287	765	510.9588	7560	16.7225	265	24.5651	310
100	50	0.6970	113	4.4376	725	2.2108	315	3.1361	460
100	100	3.6108	218	32.5960	1855	5.9332	395	8.2586	480
100	200	22.6149	316	370.1234	5145	26.1798	330	43.8539	350
200	50	2.2427	214	4.3349	605	2.1502	305	3.2779	450
200	100	3.9892	281	30.1342	1530	7.4743	345	8.7948	475
200	200	22.9626	312	342.4116	4335	23.9602	355	31.2051	495

Na figura 4 vemos o perfil de desempenho para o mesmo conjunto de instâncias. Ela também mostra que as variantes de ACC-HPE performam melhor que as variantes NEST neste conjunto de instâncias do problema de ponto-de-sela vetor-matriz.

FIGURA 3 – Perfil de desempenho para o problema de ponto-de-sela vetor-matriz



6.3 PROBLEMA QUADRÁTICO

Nesta seção comparamos as quatro variantes em instâncias do problema quadrático

$$\min_{x \in \Delta_n} \max_{y \in \Delta_m} \frac{1}{2} \|Bx\|^2 + \langle x, Ay \rangle$$

para diversos tamanhos de matrizes e diferentes valores de p . As matrizes foram geradas de modo similar às geradas para o problema bilinear (veja Subseção 6.1). A tolerância usada foi de $\bar{\varepsilon} = 10^{-4}$.

Na tabela 3 reportamos os resultados para as quatro variantes aplicados às diversas instâncias deste problema com vários tamanhos de matrizes e diferentes valores de p .

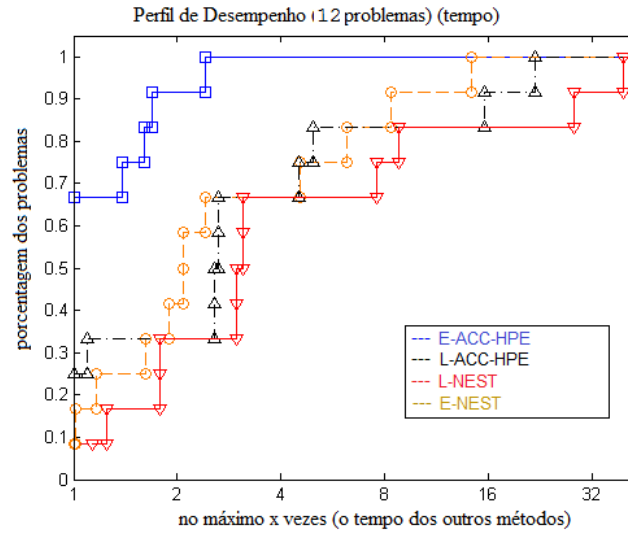
TABELA 3 – Resultados dos testes para o problema quadrático

Tamanho			E-ACC-HPE		E-Nest		L-ACC-HPE		L-Nest	
n	m	p	tempo	iter.	tempo	iter.	tempo	iter.	tempo	iter.
100	100	0.01	2.9931	3325	3.4325	3865	15.5991	16485	24.5904	29115
100	1000	0.01	1.5389	1375	22.0383	19595	33.8117	30060	42.1352	39040
1000	100	0.01	6.1160	4535	4.3215	2675	21.3346	12010	34.0820	20285
1000	1000	0.01	1.1044	595	8.8350	4960	22.6147	9230	53.5389	23255
100	100	0.1	4.4928	4985	8.6743	10345	10.9572	13125	12.0119	15415
100	1000	0.1	6.7362	5040	39.7344	31185	15.3777	12915	18.1534	15010
1000	100	0.1	42.2032	15105	41.8742	15310	47.2913	16540	52.9498	18810
1000	1000	0.1	116.7657	22820	118.2056	22850	50.3541	9450	84.0200	16820
100	100	0.2	7.7375	8585	14.9389	17815	18.8705	22605	20.6869	26545
100	1000	0.2	11.6011	8675	68.4306	39355	26.4835	22240	31.2638	25845
1000	100	0.2	28.8739	26015	72.1152	26415	81.4451	16275	91.1901	28970
1000	1000	0.2	72.6806	39305	203.5757	53710	86.7198	28490	144.6992	32390

Na figura 5 apresentamos o perfil de desempenho para o mesmo conjunto de instâncias.

Ela mostra que a variante de ACC-HPE baseada na distância Euclidiana (respectivamente, distância induzida pela divergência de Kulback-Leibler) performa melhor que a variante NEST baseada na distância Euclidiana (respectivamente, distância induzida pela divergência de Kulback-Leibler) para este conjunto de instâncias do problema quadrático.

FIGURA 4 – Perfil de desempenho para o problema quadrático



A seguir fazemos algumas considerações finais a respeito dos testes: Mostramos na Seção 5.2 que o Algoritmo 12 tem a mesma complexidade que a técnica de suavização de Nesterov de (NESTEROV, 2005). Os experimentos destas seções envolvendo três classes de problemas mostraram que as variantes do Algoritmo 12 são superiores às variantes do algoritmo de suavização de Nesterov. Os experimentos ainda mostraram que a variante do Algoritmo 12 baseada na distância Euclidiana performa melhor que a variante do Algoritmo 12 baseada na distância induzida pela divergência de Kulback-Leibler.

7 CONCLUSÃO E TRABALHOS FUTUROS

Neste capítulo fazemos considerações finais acerca de todo o trabalho e descrevemos perspectivas para desenvolvimento futuro. Nossas contribuições foram o algoritmo acelerado de HPE para ponto-de-sela (Algoritmo 5) apresentado no Capítulo 2, o algoritmo acelerado para problema de equilíbrio de Nash (Algoritmo 9), mostrado no Capítulo 3 e o algoritmo acelerado para ponto-de-sela não-Euclidiano (Algoritmo 12), apresentado no Capítulo 5. Apesar de o Algoritmo 11 aplicado para a distância Euclidiana resultar em um algoritmo Euclidiano para o problema de ponto-de-sela, este não resulta no mesmo Algoritmo 5, visto que este último não faz uso de hipótese de limitação no conjunto viável. Podemos dizer que o Algoritmo 5 é uma generalização direta do algoritmo proposto por He e Monteiro em (HE; MONTEIRO, 2014), enquanto que o caso Euclidiano do Algoritmo 11 pode ser visto como um algoritmo alternativo para resolver a mesma espécie de problemas de ponto-de-sela, sob diferentes hipóteses. Consideramos que os algoritmos propostos para os problemas de ponto-de-sela são contribuições relevantes, visto que eles atingem a complexidade ótima, conforme visto no fim da Seção 5.2. Apesar de não termos feito análise de complexidade similar para o Algoritmo 5, ela pode ser feita da mesma maneira para a mesma escolha de λ , e com isso o Algoritmo 5 também atinge a complexidade ótima para sua classe de problemas. O Algoritmo 11 foi escrito em forma de artigo e submetido para a revista Optimization Methods and Software do jornal Taylor and Francis (KOLOSSOSKI; MONTEIRO, 2015). Neste trabalho realizamos os testes relativos ao Algoritmo 11. Consideramos que tais testes foram satisfatórios comparados ao algoritmo de suavização de Nesterov.

Existem diversas maneiras de se prosseguir com o desenvolvimento deste trabalho

no futuro. Pode-se pensar em generalizar Algoritmo 9 para o contexto de distâncias de Bregman ao invés de considerar-se a Euclidiana. Pode-se fazer uma implementação numérica mais avançada aplicando-se os algoritmos desenvolvidos para problemas reais da área de economia, energia e afins. Outra perspectiva de trabalho futuro possível é a de se considerar outros problemas que possam ser reduzidos a um problema de inclusão monótona e então desenvolver um algoritmo HPE para o problema em questão. Exemplos de problemas onde isto ainda não foi feito são o problema de equilíbrio geral, que é conhecido por abranger os problemas de equilíbrio de Nash e de ponto-de-sela, problemas de minimização, de desigualdade variacional, entre outros. Outro problema que pode ser considerado é o problema de minimização multiobjetivo. Se obtido sucesso no contexto da distância Euclidiana em tais generalizações, pode-se tentar conseguir adaptar o algoritmo para o contexto de distâncias de Bregman. Pode-se ainda tentar obter resultados de complexidade pontual para o Algoritmo 12, visto que apenas resultados de complexidade ergódica foram obtidos. Outra perspectiva é considerar relaxar hipóteses, um exemplo sendo retirar a hipótese de ter que se considerar o domínio limitado no caso do Algoritmo 12.

REFERÊNCIAS

- AUSLENDER, A.; TEBoulLE, M. *Asymptotic cones and functions in optimization and variational inequalities*. New York: Springer-Verlag, 2003. xii+249 p. (Springer Monographs in Mathematics). ISBN 0-387-95520-8.
- BURACHIK, R. S.; IUSEM, A. N.; SVAITER, B. F. Enlargement of monotone operators with applications to variational inequalities. *Set-Valued Anal.*, v. 5, n. 2, p. 159–180, 1997. ISSN 0927-6947.
- BURACHIK, R. S.; SAGASTIZÁBAL, C. A.; SVAITER, B. F. ϵ -enlargements of maximal monotone operators: theory and applications. In: *Reformulation: nonsmooth, piecewise smooth, semismooth and smoothing methods (Lausanne, 1997)*. Dordrecht: Kluwer Acad. Publ., 1999, (Appl. Optim., v. 22). p. 25–43.
- CENSOR, Y.; ZENIOS, S. A. Proximal minimization algorithm with D-functions. *Journal of Optimization Theory and Applications*, v. 73, 1992.
- CHEN, G.; TEBoulLE, M. Convergence analysis of a proximal-like minimization algorithm using bregman functions. *SIAM Journal on Optimization*, v. 3, p. 538–543, 1993.
- CHEN, Y.; LAN, G.; OUYANG, Y. Accelerated schemes for a class of variational inequalities. *submitted to Mathematical Programming*, 2014.
- DANG, C. D.; LAN, G. On the convergence properties of non-euclidean extragradient methods for variational inequalities with generalized monotone operators. *Computational Optimization and applications*, v. 60, n. 2, p. 277–310, 2015.
- DOLAN, E. D.; MORE, J. J. Benchmarking optimization software with performance profiles. *Mathematical Programming*, v. 91, n. 2, p. 201–213, 2002.
- ECKSTEIN, J. Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming. *Mathematics of Operations Research*, v. 18, n. 1, 1993.
- EGGERMONT, P. P. B. Multiplicative iterative algorithms for convex programming. *Linear Algebra and Applications*, v. 130, p. 25–32, 1990.
- HE, Y.; MONTEIRO, R. D. C. Accelerating block-decomposition first-order methods for solving generalized saddle-point and Nash equilibrium problems. *Optimization-online preprint*, 2013.
- HE, Y.; MONTEIRO, R. D. C. An accelerated HPE-type algorithm for a class of composite convex-concave saddle-point problems. *working paper*, 2014.
- IUSEM, A. N.; SOLODOV, M. V. Newton-type methods with generalized distances for constrained optimization. *Optimization*, v. 41, p. 257–278, 1997.
- KIWIEL, K. Proximal minimization methods with generalized Bregman functions. *SIAM Journal on Control Optimization*, v. 35, p. 1142–1168, 1997.

- KOLOSSOSKI, O.; MONTEIRO, R. D. C. An accelerated non-euclidean hybrid proximal extragradient-type algorithm for convex-concave saddle-point problems. *submitted to Optimizations Methods and Software*, 2015.
- KORPELEVIČ, G. M. An extragradient method for finding saddle points and for other problems. *Èkonom. i Mat. Metody*, v. 12, n. 4, p. 747–756, 1976. ISSN 0424-7388.
- MINTY, G. Monotone nonlinear operators in hilbert space. *Duke Mathematical Journal*, v. 29, p. 341–346, 1978.
- MONTEIRO, R. D. C.; SVAITER, B. Complexity of variants of Tseng’s modified F-B splitting and korpelevich’s methods for hemivariational inequalities with applications to saddle-point and convex optimization problems. *SIAM Journal on Optimization*, SIAM, v. 21, n. 4, p. 1688–1720, 2011.
- MONTEIRO, R. D. C.; SVAITER, B. F. Convergence rate of inexact proximal point methods with relative error criteria for convex optimization. *submitted to SIAM Journal on Optimization*, 2010.
- MONTEIRO, R. D. C.; SVAITER, B. F. On the complexity of the hybrid proximal extragradient method for the iterates and the ergodic mean. *SIAM Journal on Optimization*, SIAM, v. 20, n. 6, p. 2755–2787, 2010.
- MONTEIRO, R. D. C.; SVAITER, B. F. Iteration-complexity of block-decomposition algorithms and the alternating direction method of multipliers. *SIAM Journal on Optimization*, SIAM, v. 23, n. 1, p. 475–507, 2013.
- NEMIROVSKI, A. Prox-method with rate of convergence $O(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems. *SIAM Journal on Optimization*, SIAM, v. 15, n. 1, p. 229–251, 2004.
- NEMIROVSKY, A. Information-based complexity of linear operator equations. *Journal of Complexity*, v. 8, p. 153–175, 1992.
- NESTEROV, Y. *Introductory Lectures on Convex Optimization: A Basic Course*. [S.l.]: Kluwer Academic Publishers, 2004.
- NESTEROV, Y. Smooth minimization of non-smooth functions. *Mathematical Programming*, Springer, v. 103, n. 1, p. 127–152, 2005.
- ROCKAFELLAR, R. T. Characterization of the subdifferentials of convex functions. *Pacific J. Math.*, v. 17, p. 497–510, 1966.
- ROCKAFELLAR, R. T. Convex analysys. *Princeton Math. Series*, Princeton Univ. Press, v. 28, 1970.
- ROCKAFELLAR, R. T. On the maximal monotonicity of subdifferential mappings. *Pacific J. Math.*, v. 33, p. 209–216, 1970. ISSN 0030-8730.
- ROCKAFELLAR, R. T. On the maximality of sums of nonlinear monotone operators. *Trans. Amer. Math. Soc.*, v. 149, p. 75–88, 1970. ISSN 0002-9947.

ROCKAFELLAR, R. T. On the virtual convexity of the domain and range of a nonlinear maximal monotone operator. *Math. Annalen*, v. 185, p. 81–90, 1970.

SOLODOV, M. V.; SVAITER, B. F. A hybrid approximate extragradient-proximal point algorithm using the enlargement of a maximal monotone operator. *Set-Valued Anal.*, v. 7, n. 4, p. 323–345, 1999. ISSN 0927-6947.

SOLODOV, M. V.; SVAITER, B. F. A hybrid projection-proximal point algorithm. *J. Convex Anal.*, v. 6, n. 1, p. 59–70, 1999. ISSN 0944-6532.

SOLODOV, M. V.; SVAITER, B. F. An inexact hybrid generalized proximal point algorithm and some new results on the theory of Bregman functions. *Math. Oper. Res.*, v. 25, n. 2, p. 214–230, 2000. ISSN 0364-765X.

SOLODOV, M. V.; SVAITER, B. F. A unified framework for some inexact proximal point algorithms. *Numer. Funct. Anal. Optim.*, v. 22, n. 7-8, p. 1013–1035, 2001. ISSN 0163-0563.

TSENG, P. A modified forward-backward splitting method for maximal monotone mappings. *SIAM J. Control Optim.*, v. 38, n. 2, p. 431–446 (electronic), 2000. ISSN 0363-0129. Disponível em: <http://dx.doi.org/10.1137/S0363012998338806>.

TSENG, P. On accelerated proximal gradient methods for convex-concave optimization. *submitted to Journal on Optimization*, 2008.